

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

CODIFICADORES CELP: IMPLEMENTAÇÃO, ANÁLISE E
PROPOSTAS PARA REDUÇÃO DE COMPLEXIDADE E TAXA DE TRANSMISSÃO

DISSERTAÇÃO SUBMETIDA À UNIVERSIDADE FEDERAL DE SANTA CATARINA
PARA OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA

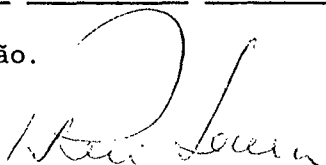
DÊNIS FERNANDES

FLORIANÓPOLIS, DEZEMBRO DE 1992.

**CODIFICADORES CELP: IMPLEMENTAÇÃO, ANÁLISE E
PROPOSTAS PARA REDUÇÃO DE COMPLEXIDADE E TAXA DE TRANSMISSÃO**

CANDIDATO: DÊNIS FERNANDES

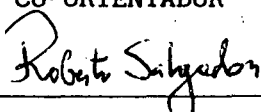
Esta dissertação foi julgada para obtenção do título MESTRE EM ENGENHARIA - ESPECIALIDADE ENGENHARIA ELÉTRICA e aprovada em sua forma final pelo Curso de Pós-Graduação.



Prof. Rui Seara, Dr. Ing.
ORIENTADOR



Prof. José Carlos Moreira Bermudez, Ph. D.
CO-ORIENTADOR

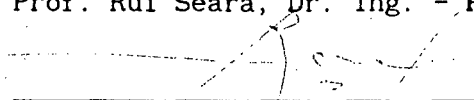


Prof. Roberto de Souza Salgado, Ph. D.
COORDENADOR DO CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

BANCA EXAMINADORA:



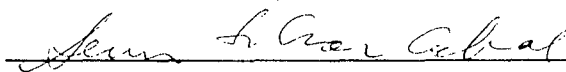
Prof. Rui Seara, Dr. Ing. - **PRESIDENTE**



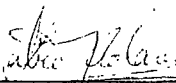
Prof. José Carlos Moreira Bermudez, Ph. D.



Prof. Hans Helmut Zürn, Ph. D.



Prof.^a Leonor Scliar-Cabral, Dr.



Prof. Fábio Violaro, Dr.

*Na profundidade do inconsciente humano,
existe uma necessidade penetrante de um
universo lógico, que faça sentido. Mas o
universo real está sempre um passo
adiante da lógica.*

Duna de Frank Herbert

*As pessoas que acreditam que a vontade
suplanta suas limitações.*

As pessoas que acreditam no trabalho.

As pessoas sinceras e honestas.

*As pessoas que não têm medo de fazer o
que a alma deseja.*

Em especial, à Crismara Janina.

I N D I C E

LISTA DE FIGURAS.....	ix
LISTA DE TABELAS.....	xi
SIMBOLOGIA.....	xv
RESUMO.....	xviii
ABSTRACT.....	xix
 <u>CAPÍTULO 1</u> - INTRODUÇÃO.....	 1
 <u>CAPÍTULO 2</u> - CODIFICAÇÃO DE SINAIS DE VOZ.....	 8
2.1 - Introdução.....	8
2.2 - Codificação da Forma de Onda.....	10
2.3 - Codificação Paramétrica.....	13
2.3.1 - Modelo Básico para Síntese de Voz.....	14
2.3.2 - Vocoders.....	21
2.4 - Conclusões.....	28
 <u>CAPÍTULO 3</u> - CODIFICAÇÃO CELP.....	 30
3.1 - Introdução.....	30
3.2 - Determinação dos Parâmetros do Codificador CELP Convencional.....	34
3.2.1 - Determinação dos Coeficientes do Filtro PCA.....	34
3.2.2 - Determinação dos Parâmetros do Filtro PLA.....	39
3.2.3 - Determinação do Ganho e do Índice da Codeword.....	44
3.2.4 - Ponderação do Erro.....	45
3.2.5 - Expansão da Banda do Filtro PCA.....	46

3.2.6 - Pós-Filtragem.....	47
3.2.7 - Ceifamento Central do "Codebook".....	49
3.3 - Quantização dos Parâmetros do Codificador CELP.....	49
3.3.1 - Quantização dos Coeficientes do Filtro PCA.....	49
3.3.2 - Quantização dos Parâmetros do Filtro PLA.....	52
3.3.3 - Quantização do Ganho.....	52
3.3.4 - Número de "Codewords".....	53
3.4 - Análise do Desempenho do Codificador CELP.....	53
3.4.1 - Variação da Ordem do Filtro PCA.....	55
3.4.2 - Variações no Janelamento do Sinal de Voz.....	56
3.4.3 - Utilização da Memória Inicial dos Filtros PCA e Ponderador de Erro.....	58
3.4.4 - Método de Determinação dos Coeficientes do Filtro PCA.....	59
3.4.5 - Variação do Intervalo de Atualização dos Coeficientes do Filtro PCA.....	61
3.4.6 - Método de Determinação do Retardo do Filtro PLA.....	62
3.4.7 - Variação da Faixa do Retardo do Filtro PLA.....	63
3.4.8 - Método de Determinação dos Coeficientes do Filtro PLA.....	64
3.4.9 - Variação do Número de Coeficientes do Filtro PLA.....	65
3.4.10 - Estabilização do Filtro PLA.....	65
3.4.11 - Ceifamento Central do "Codebook".....	67
3.4.12 - Variação da Duração do Subquadros.....	68
3.4.13 - Quantização dos Parâmetros.....	69
3.4.14 - Desempenho em Taxas Diversas.....	82
3.4.15 - Variação do Fator de Ponderação do Erro.....	83
3.4.16 - Variação do Fator de Expansão da Banda do Filtro PCA.....	84
3.4.17 - Pós-Filtragem.....	84

3.4.18 - Variação do Número de "Codewords" para as Taxas Base de 7200 bps e 9600 bps.....	86
3.5 - Conclusões.....	87
 <u>CAPÍTULO 4</u> - CODIFICAÇÃO CMSEC.....	89
4.1 - Introdução.....	89
4.2 - Determinação dos Parâmetros do Codificador CMSEC.....	89
4.3 - Quantização dos Parâmetros do Codificador CMSEC.....	93
4.4 - Análise do Desempenho do Codificador CMSEC.....	94
4.5 - Complexidade Computacional dos Codificadores CELP e CMSEC.....	102
4.6 - Conclusões.....	106
 <u>CAPÍTULO 5</u> - QUANTIZAÇÃO ESTATÍSTICA DO GANHO.....	108
5.1 - Introdução.....	108
5.2 - Procedimento de Quantização Estatística do Ganho.....	109
5.3 - Análise do Desempenho dos Codificadores CELP e CMSEC com Quantização Estatística dos Ganhos.....	113
5.4 - Conclusões.....	117
 <u>CAPÍTULO 6</u> - DETERMINAÇÃO RÁPIDA DO RETARDO DO FILTRO PLA.....	119
6.1 - Introdução.....	119
6.2 - Procedimento de Busca Rápida de M.....	120
6.3 - Análise do Desempenho dos Codificadores CELP e CMSEC com o Procedimento de Busca Rápida de M.....	123
6.4 - Conclusões.....	126
 <u>CAPÍTULO 7</u> - CONCLUSÕES.....	127

REFERÊNCIAS BIBLIOGRÁFICAS.....	131
<u>APÊNDICE A</u> - MÉTODO DE BLANKINSHIP PARA O CÁLCULO DE AUTOCORRELAÇÕES.....	150
<u>APÊNDICE B</u> - MÉTODO DE CHOLSKY PARA RESOLUÇÃO DE SISTEMAS LINEARES.....	154
<u>APÊNDICE C</u> - ARQUIVOS COM OS SEGMENTOS DE VOZ UTILIZADOS NAS SIMULAÇÕES....	157
<u>APÊNDICE D</u> - TEMPO DE PROCESSAMENTO PARA O CODIFICADOR CELP IMPLEMENTADO...	160

LISTA DE FIGURAS

2.1 - Codificação de sinais de voz.....	8
2.2 - Taxas de transmissão em função do tipo de codificação.....	10
2.3 - Modelo físico aproximado do sistema vocal humano.....	14
2.4 - Forma de onda e densidade espectral de potência da vogal "a".....	15
2.5 - Forma de onda e densidade espectral de potência da consoante "s".....	16
2.6 - Modelo de tubos para produção de sons nasais.....	17
2.7 - Modelo fonte/filtro para produção de voz.....	18
2.8 - Modelo para síntese de voz.....	19
2.9 - Forma aproximada da excitação glotal.....	20
2.10 - Vocoder de canal (codificador).....	22
2.11 - Canal de um vocoder de fase (codificador).....	23
2.12 - Vocoder homomórfico (codificador).....	24
2.13 - Vocoder de formantes (codificador).....	25
2.14 - Vocoder LPC (codificador).....	26
2.15 - Vocoder LPC excitado por voz.....	27
3.1 - Modelo de análise do codificador CELP convencional.....	30
3.2 - Modelo de síntese do codificador CELP convencional.....	31
3.3 - Procedimento de segmentação e ponderação do sinal de voz para análise LPC.....	37
3.4 - Resposta em frequência de um filtro PCA para $\delta = 1$ e $\delta = 0,95$	47
3.5 - Sistema de pós-filtragem combinado.....	48
3.6 - Relação sinal/ruído segmentada em função da ordem do filtro PCA.....	56
3.7 - Segmentos de voz obtidos com os métodos da autocorrelação e covariância.....	60-1

3.8 - Segmentos de voz obtidos com filtros PLA não instáveis e instáveis..	66-7
3.9 - Histogramas das razões log-área.....	70-3
3.10 - Histograma do $\ln b_0 $	76
4.1 - Modelo de análise do codificador CMSEC.....	90
4.2 - Modelo de síntese do codificador CMSEC.....	91
4.3 - Ocorrências de $\ln g_1 $ no codificador CMSEC de 3 ganhos.....	97-8
5.1 - Histograma do $\ln g $ no codificador CELP e densidade de probabilidade normal com mesmas média e variância.....	109
6.1 - Função $E^{-1}(M)$ para um segmento vozeado.....	121
6.2 - Função $E^{-1}(M)$ para um segmento não vozeado.....	121
6.3 - Procedimento de busca rápida de M.....	122
A.1 - Número de multiplicações no cálculo de $R(\tau)$	152

LISTA DE TABELAS

3.1 - Número de bits e limites para quantização do seno inverso dos PARCOR.....	51
3.2 - Parâmetros adotados como referência para o codificador CELP.....	54
3.3 - Resultados objetivos obtidos como referência inicial para o codificador CELP.....	54
3.4 - Resultados objetivos em função da variação da ordem do filtro PCA no codificador CELP.....	55
3.5 - Resultados objetivos em função do janelamento do sinal de voz no codificador CELP.....	57
3.6 - Resultados objetivos com utilização ou não da memória inicial nos filtros PCA e ponderador do erro.....	58
3.7 - Resultados objetivos com diversos métodos para determinação dos coeficientes do filtro PCA.....	59
3.8 - Resultados objetivos em função da variação do quadro de análise.....	62
3.9 - Resultados objetivos em função do método de determinação do retardo do filtro PLA.....	63
3.10 - Resultados objetivos em função da faixa do retardo do filtro PLA.....	63
3.11 - Resultados objetivos em função do método de determinação dos coeficientes do filtro PLA.....	64
3.12 - Resultados objetivos em função do número de coeficientes do filtro PLA.....	65
3.13 - Resultados objetivos obtidos com a limitação de $ b_0 $	66
3.14 - Resultados objetivos para vários níveis de ceifamento central do "codebook".....	67

3.15 - Resultados objetivos em função do número de amostras do subquadro....	68
3.16 - Limites de quantização para o seno inverso dos PARCOR e razões log- área.....	70
3.17 - Resultados objetivos com a limitação e quantização infinita dos parâmetros do filtro PCA.....	74
3.18 - Bits para quantização dos parâmetros do filtro PCA.....	74
3.19 - Resultados objetivos com a quantização do seno inverso dos PARCOR e das razões log-área.....	74
3.20 - Resultados objetivos em função do número de bits na quantização das razões log-área.....	75
3.21 - Limites para quantização do coeficiente do filtro PLA.....	76
3.22 - Resultados objetivos com a limitação de b_0 na quantização.....	77
3.23 - Resultados objetivos em função do número de bits na quantização de b_0	77
3.24 - Limites para quantização de $\ln g $	78
3.25 - Resultados objetivos em função da limitação do ganho.....	78
3.26 - Resultados objetivos em função do número de bits na quantização do ganho.....	79
3.27 - Resultados objetivos em função do número de "codewords".....	80
3.28 - Distribuição de bits e taxas por etapa para o codificador CELP de 7200 bps.....	81
3.29 - Desempenho objetivo do codificador CELP com taxas de 6800 bps e 7200 bps.....	82
3.30 - Resultados objetivos com taxas entre 9600 e 2400 bps.....	82
3.31 - Resultados objetivos em função da variação do fator de ponderação do erro em 7200 bps.....	83
3.32 - Resultados objetivos em função do fator de expansão da banda do	

filtro PCA em 7200 bps.....	84
3.33 - Resultados objetivos em 4800 bps com pós-filtragem.....	85
3.34 - Resultados objetivos em 3600 e 2400 bps com utilização de pós-filtragem.....	86
3.35 - Resultados objetivos com a redução do número de "codewords" a partir da configuração básica de 9600 bps.....	86
3.36 - Resultados objetivos com a redução do número de "codewords" a partir da configuração básica de 7200 bps.....	87
4.1 - Resultados objetivos para o codificador CMSEC com quantização infinita dos ganhos.....	95
4.2 - Limites para quantização dos ganhos no codificador CMSEC obtidos estatisticamente.....	97
4.3 - Resultados objetivos obtidos com o codificador CMSEC de 3 ganhos e 64 "codewords".....	99
4.4 - Resultados objetivos obtidos com o codificador CMSEC de 2 ganhos e 256 "codewords".....	99
4.5 - Resultados objetivos obtidos com o codificador CMSEC de 8 ganhos e 2 "codewords".....	100
4.6 - Resultados objetivos obtidos com o codificador CMSEC de 5 ganhos e 128 "codewords".....	100
4.7 - Resultados objetivos do codificador CMSEC a 4800 bps.....	101
4.8 - Operações aritméticas para determinação da excitação no codificador CELP convencional operando a 7200 bps e 9600 bps.....	104
4.9 - Operações aritméticas para determinação da excitação no codificador CMSEC de 5 ganhos e 128 "codewords".....	105
4.10 - Operações aritméticas para determinação da excitação no codificador CMSEC de 2 ganhos e 256 "codewords".....	105

4.11 - Operações aritméticas para determinação da excitação no codificador CMSEC de 3 ganhos e 64 "codewords".....	106
4.12 - Operações aritméticas para determinação da excitação no codificador CMSEC de 8 ganhos e 2 "codewords".....	106
5.1 - Valores utilizados para quantização estatística do ganho no codificador CELP.....	114
5.2 - Resultados objetivos em função do procedimento de quantização do ganho no codificador CELP com 1024 "codewords".....	114
5.3 - Valores utilizados para quantização estatística dos ganhos nos codificadores CMSEC de 2, 3, 5 e 8 ganhos.....	115
5.4 - Resultados objetivos com a aplicação de quantização estatística dos ganhos no codificador CMSEC.....	116
5.5 - Resultados objetivos para o codificador CMSEC a 4800 bps com quantização estatística dos ganhos.....	117
6.1 - Resultados objetivos em função de diversos Δ_0 para o codificador CELP a 7200 bps.....	123
6.2 - Resultados objetivos para o codificador CMSEC a 10800 bps com o procedimento de busca rápida de M.....	124
6.3 - Resultados objetivos para o codificador CELP a 7200 bps, com $\Delta_0 = 1$ e $\Delta_0 = 8$	125
6.4 - Resultados objetivos para os codificadores CELP e CMSEC a 9600 bps com quantização estatística e busca rápida de M.....	125
C.1 - Arquivos com os segmentos de voz utilizados nas simulações.....	159
D.1 - Tempos de processamento do programa de simulação do codificador CELP convencional em 7200 bps para diversas máquinas.....	161

S I M B O L O G I A

- a_1 - coeficientes do filtro preditor de curto atraso
 b_j - coeficientes do filtro preditor de longo atraso
 bps - bits por segundo
 $c_k(n)$ - k-ésima "codeword" no codificador CELP
 $c_{1k}(n)$ - k-ésima "codeword" do l-ésimo "codebook" no codificador CMSEC
 CMSEC - "Constrained Multiple-Stage Excitation Coding"
 $e(n)$ - erro não ponderado
 $\hat{e}(n)$ - erro ponderado
 E - energia de uma seqüência
 $E(M)$ - energia do erro em função de M
 $E^{-1}(M)$ - inverso da energia do erro em função de M
 $f(n)$ - resposta ao impulso do filtro preditor de curto atraso
 f_s - frequência de amostragem
 $F(z)$ - Função de transferência do filtro preditor de curto atraso
 g - ganho do codificador CELP convencional
 g_1 - l-ésimo ganho do codificador CMSEC
 K - número de "codewords" no "codebook"
 k_i - i-ésimo coeficiente PARCOR
 K_i - i-ésima razão log-área
 $l(n)$ - resposta ao impulso do filtro preditor de longo atraso
 L - número de ganhos ou "codebooks" na codificador CMSEC
 $L(z)$ - Função de transferência do filtro preditor de longo atraso
 M - retardo do preditor de longo atraso
 N - número de amostras do subquadro

N_q - número de amostras do quadro

OP - ordem do preditor de curto atraso

P - número de termos da série no método de quantização estatística do ganho

$p(n)$ - resposta ao impulso do filtro ponderador de erro

$p_0(n)$ - resposta à entrada zero do filtro ponderador de erro

$P(z)$ - função de transferência do filtro ponderador de erro

PCA - Preditor de Curto Atraso

Pitch - período fundamental do sinal de voz, relacionado com o timbre

PLA - Preditor de Longo Atraso

$r(n)$ - resíduo do filtro PCA

$R(\tau)$ - função autocorrelação no ponto τ

RSRSEG - Relação Sinal/Ruído SEGmentada (dB)

$s(n)$ - amostras do sinal de voz original

$\hat{s}(n)$ - sinal de voz menos a resposta à entrada zero do filtro PCA, ponderados

$\tilde{s}_0(n)$ - resposta à entrada zero da cascata de filtros PLA e PCA

$\tilde{s}(n)$ - sinal de voz sintetizado menos a resposta à entrada zero da cascata de
filtros PLA e PCA

$t(n)$ - resíduo do filtro PLA

T - número de coeficientes do filtro PLA

T_p - tempo de processamento (minutos)

$x(n)$ - excitação do filtro PLA

$y(n)$ - saída do filtro PLA

$z(n)$ - saída do filtro PCA ponderado, com condições iniciais nulas e excitado
pela resposta à entrada zero do filtro PLA

α, β - fatores de pós-filtragem do filtro PCA

δ - fator de expansão da banda do filtro PCA

Δf - expansão da banda do filtro PCA (Hz)

Δ_0 - passo inicial no procedimento de busca rápida de M

ϵ - fator de pós-filtragem do filtro PLA

γ - fator de ponderação do erro

$\Phi_{sy}(i, j)$ - função correlação cruzada de $s(n)$ e $y(n)$

μ - média

σ - desvio padrão

R E S U M O

O codificador CELP ("Code-Excited Linear Predictive") é provavelmente a estrutura de codificação de voz mais estudada atualmente. O sucesso deste modelo deve-se basicamente à boa qualidade do sinal de voz sintetizado em taxas tão baixas quanto 4800 bps, o que não tem sido possível com outras formas de codificação. O principal problema do codificador CELP diz respeito à determinação da seqüência de excitação, que é normalmente escolhida através de uma busca exaustiva em um conjunto de seqüências denominado "codebook".

Neste trabalho, é estudado o comportamento do codificador CELP convencional frente a variações em seus parâmetros, com o objetivo de estabelecer padrões de referência para comparação com novas estruturas de codificação que operem em taxas abaixo de 10 kbps. São apresentadas ainda, três propostas para diminuição de complexidade e aumento de qualidade do sinal de voz sintetizado. A primeira proposta consiste de uma estrutura de codificação paramétrica com complexidade reduzida em relação ao CELP convencional. A segunda proposta é um procedimento de quantização do ganho que leva em conta uma aproximação da função densidade de probabilidade do logarítmo do módulo deste parâmetro. A terceira e última proposta é um método de redução de complexidade na determinação do retardo do filtro preditor de longo atraso.

O desempenho das diversas configurações de codificadores testadas é avaliado quanto à qualidade do sinal sintetizado, em termos objetivos (relação sinal/ruído segmentada) e subjetivos (análise comparativa informal).

ABSTRACT

The CELP (Code-Excited Linear Predictive) is probably the most studied speech coder at the present. The success of this model is originated from the good quality of the synthesized speech signals at rates as low as 4800 bps, what is not possible with others coding procedures. The main problem of the CELP coder is determining the excitation sequence, normally picked out through an exhaustive search in a set called codebook.

In this work the performance of the conventional CELP coder is studied at several different parameters configurations, which is aimed to determine standard configurations to compare the performance of new coding models operating at rates below 10 kbps. It is also presented, three propositions to complexity reduction and to increase the quality of the synthesized voice signal. The first proposition is a model of parametric speech coder with reduced computational complexity in confront with the conventional CELP coder. The second proposition is a gain quantization procedure which considers an approximation of the probability density function of the natural logarithm of the absolute gain parameter. The third is a method to reduce the computational complexity in determining the delay of the long delay prediction filter.

The performance of the several analyzed coders configurations is compared regarding the quality of the synthesized speech signal in both, objective (segmental signal-to-noise ratio) and subjective senses (informal comparative analysis).

C A P Í T U L O 1

I N T R O D U Ç Ã O

A voz é um dos instrumentos de comunicação mais desenvolvidos no ser humano. O processo de fala carrega em si uma grande quantidade de informações codificadas sob a forma de linguagem, que são transmitidas como ondas acústicas obtidas através de um processo de modulação de pressão no meio de transmissão.

Os sinais de voz são compostos de uma seqüência de sons representados por símbolos denominados fonemas, cuja disposição é regida por regras de linguagem. O conjunto de fonemas é uma das características de cada linguagem, possuindo normalmente entre 30 e 50 fonemas [1]. O estudo das regras de linguagem e suas implicações na comunicação humana é denominado lingüística, enquanto que o estudo dos sons da voz é denominado fonética.

A necessidade do tratamento de informações na forma de sinais de voz, tais como a transmissão a grandes distâncias ou o armazenamento de tais informações para posterior utilização, implica transformação dos sinais acústicos em sinais de outra natureza, de forma que o procedimento desejado seja mais facilmente implementado. A conversão de sinais acústicos em sinais elétricos é normalmente utilizada dada a multiplicidade de dispositivos disponíveis para manipulação de sinais desta natureza.

Os sinais elétricos utilizados para representar um sinal de voz podem ser analógicos, onde a amplitude de uma grandeza elétrica, tal como tensão ou corrente, segue a amplitude da pressão acústica, ou numéricos, onde a

amplitude da pressão acústica é representada por um determinado código numérico, correspondente à amostragem do sinal em intervalos regulares de tempo.

O avanço na área das técnicas digitais tem possibilitado a implementação econômica de dispositivos com desempenho tão bom ou até melhor sob certos aspectos do que seus equivalentes analógicos. Aliados a este fato, a simplicidade de implementação e o aprimoramento de métodos específicos têm proporcionado uma grande evolução das técnicas de processamento digital de sinais de voz.

O estudo de sinais de voz tem encontrado interesse em diversas aplicações que podem ser incluídas nos grandes grupos relacionados a seguir:

- reconhecimento de voz:
 - reconhecimento de locutor
 - reconhecimento de locução
- síntese a partir de texto escrito
- regeneração de sinais de voz:
 - redução de ruído
 - cancelamento de ecos
 - etc.
- codificação de voz:
 - compressão para armazenamento
 - redução da taxa de transmissão

Os sistemas de reconhecimento de voz distinguem o locutor e/ou a locução. Assim, um sistema de reconhecimento de voz pode identificar uma determinada pessoa através da análise de suas características vocais ou ainda identificar uma palavra ou mensagem oriunda de um locutor qualquer. Estas identificações habilitam o sistema a executar qualquer tarefa

pré-estabelecida. Os sistemas de reconhecimento de locutor podem ser ainda classificados em sistemas de verificação de locutor e sistemas de identificação de locutor. No primeiro, o sistema simplesmente aceita ou rejeita um locutor segundo sua identidade aclamada. No segundo, o sistema identifica o locutor entre uma série de possibilidades através da comparação de padrões [2,3,4].

No processo de síntese de voz a partir de um texto escrito, o sistema analisa o texto a ser lido e sintetiza a mensagem de voz correspondente através da aplicação de regras adequadas de pronúncia e entonação. Sistemas desta natureza podem ser utilizados, por exemplo, no auxílio a portadores de deficiências visuais [5,6,7].

A regeneração do sinal de voz é necessária quando ocorre uma degradação da qualidade do mesmo. Desta forma, podem ser atenuados os ecos, os ruídos e outras perturbações introduzidas no sinal original durante sua geração, transmissão ou armazenamento, melhorando sua qualidade final.

O armazenamento de sinais de voz tem aplicações nas chamadas máquinas falantes, onde uma mensagem original é armazenada de forma compactada, possibilitando à máquina sintetizar a mesma quando necessário. A redução na taxa de bits do sinal de voz original implica diretamente em menor quantidade de memória para o armazenamento das informações [8].

A redução da taxa de transmissão tem como objetivo essencial o maior aproveitamento dos meios de transmissão disponíveis, tais como os canais de comunicação via satélite e sistemas de telefonia móvel, por exemplo. Uma das características que distinguem os procedimentos de redução de taxa de transmissão dos procedimentos de redução de taxa para armazenamento é a necessidade de processamento em tempo real, o que implica desenvolvimento de procedimentos rápidos para codificação e decodificação dos sinais de voz.

Um problema importante com respeito à transmissão e armazenamento de sinais de voz é a velocidade com a qual a informação é convertida em ondas acústicas. Uma estimativa grosseira pode ser obtida levando-se em conta que a limitação física na articulação da voz humana implica a conversão da informação a uma taxa de aproximadamente 10 fonemas por segundo. Considerando um conjunto médio de 50 fonemas, é necessária a representação de cada um deles através de um código de 6 bits. Para a taxa média de 10 fonemas por segundo, pode ser estimado um limite mínimo em torno de 60 bps para transmissão da informação de voz [1]. Cabe ressaltar que esta estimativa considera apenas as informações contidas na locução, desconsiderando o aspecto da identificação do locutor, da entonação e do conteúdo emocional, o que certamente conduziria a uma taxa mais elevada.

Uma das mais importantes aplicações do processamento de voz é o vocoder ("voice"+"coder") inventado por Homer Dudley nos anos 30 e cuja finalidade é reduzir a largura de banda necessária para transmitir o sinal de voz. A necessidade de redução da largura de banda, determinada pela redução da taxa de bits no caso da transmissão digital, é um fator econômico que tem implicado considerável quantidade de pesquisa nesta área.

Os chamados codificadores paramétricos são exemplos típicos de aplicações que visam diminuir a taxa de transmissão ou a memória necessária para o armazenamento dos sinais de voz, aproveitando-se de características específicas destes sinais, tais como composição espectral por exemplo. Para baixas taxas, os codificadores paramétricos têm se mostrado potencialmente melhores com relação à qualidade da voz sintetizada do que os codificadores de forma de onda (PCM - "Pulse Code Modulation", ADPCM - "Adaptive Differential Pulse Code Modulation", etc.).

As redes de telefonia pública digital foram implantadas tendo como base

canais com 64 kbps que possibilitam uma única conversação telefônica. A redução da taxa de transmissão dos sinais de voz permite hoje que sejam transmitidas várias conversações telefônicas em um único canal de 64 kbps, sem perda sensível de qualidade. O CCITT ("International Telegraph and Telephone Consultative Committee") tem procedimentos padronizados para transmissão de sinais de voz em 64 kbps (recomendação G711: log-PCM lei μ ou lei A), 32 kbps (recomendação G721: ADPCM) [8] e 16 kbps (LD-CELP - "Low Delay Code-Excited Linear Predictive"), todos eles com requisitos de qualidade compatíveis com aqueles exigidos pelos sistemas de telefonia pública.

Dentre as estruturas que têm como objetivo a redução da taxa de transmissão para aplicação em sistemas de telefonia, o codificador CELP ("Code-Excited Linear Predictive") é atualmente aquela onde tem sido concentrada grande parte das pesquisas, visto a boa qualidade da voz sintetizada obtida em taxas tão baixas quanto 4800 bps. O foco principal das pesquisas está dirigido à redução da complexidade visando a aplicação em tempo real, requisito essencial para aplicações em telefonia pública, e também para a possibilidade de maior redução da taxa sem degradação significativa do sinal de voz.

No presente trabalho, é estudado o comportamento de diversas configurações do codificador CELP convencional, com o objetivo principal de estabelecer padrões de referência para comparação de novas estruturas de codificadores que operem em taxas abaixo de 10 kbps. São propostos ainda uma estrutura alternativa com menor complexidade do que o codificador CELP convencional, um procedimento de quantização do parâmetro ganho, que possibilita um aumento da qualidade do sinal de voz sintetizado sem aumento da taxa de transmissão, e um procedimento de determinação do retardo do filtro PLA com complexidade menor do que a metade daquela obtida com o procedimento

convencional.

No capítulo 2 são apresentados, de forma sucinta, diversos métodos de codificação de voz, sendo dada maior ênfase aos métodos de codificação paramétrica, onde inclui-se o codificador CELP. É também apresentado um modelo para síntese de voz que é a base de uma grande classe de codificadores paramétricos.

No capítulo 3 é apresentada a estrutura CELP convencional. São estudadas diversas formas de determinação dos parâmetros do codificador, a quantização escalar destes parâmetros e a análise do desempenho de várias configurações do codificador CELP.

No capítulo 4 é apresentada uma estrutura alternativa de codificador paramétrico com menor complexidade do que o CELP convencional, denominada CMSEC ("Constrained Multiple-Stage Excitation Coding"), na qual a excitação é formada como combinação linear de diversas seqüências aleatórias. É estudado o desempenho de diversas configurações da nova estrutura utilizando quantização escalar dos parâmetros.

No capítulo 5 é apresentado um procedimento de quantização do parâmetro ganho, denominado quantização estatística do ganho, que leva em conta a função densidade de probabilidade do parâmetro na alocação dos níveis de quantização. O procedimento proposto é aplicado nos codificadores CELP e CMSEC e os resultados são comparados com aqueles obtidos com o conhecido procedimento de quantização uniforme do logaritmo do módulo do ganho.

No capítulo 6 é apresentada uma nova forma de determinação do retardo do filtro preditor de longo atraso (PLA). Esta técnica reduz a complexidade média da determinação do retardo em mais de 50%, sem reduzir significativamente a qualidade do sinal sintetizado. São apresentados os resultados obtidos com o procedimento proposto aplicado aos codificadores CELP e CMSEC.

No Apêndice A é apresentada uma formalização de um método para o cálculo de autocorrelações com menor complexidade, e que foi utilizado nas simulações executadas.

No Apêndice B é apresentado o método de Cholesky para resolução, com complexidade computacional reduzida, de sistemas lineares cuja matriz dos coeficientes apresente a propriedade da simetria.

No Apêndice C são descritos os segmentos de voz utilizados nas simulações e nos levantamentos estatísticos.

No Apêndice D são apresentados dados relativos ao tempo de processamento requerido pelo codificador CELP de 7200 bps implementado em máquinas distintas.

CAPÍTULO 2

CODIFICAÇÃO DE SINAIS DE VOZ2.1 - INTRODUÇÃO

Existem várias técnicas para codificação digital de sinais de voz cujos resultados finais diferenciam-se basicamente quanto à qualidade da voz sintetizada, à taxa final e à complexidade computacional. Conforme apresentado na Fig.2.1, as técnicas consagradas para codificação de sinais de voz podem ser classificadas em dois grandes grupos: a codificação da forma de onda e a codificação paramétrica [1,2,9,10].

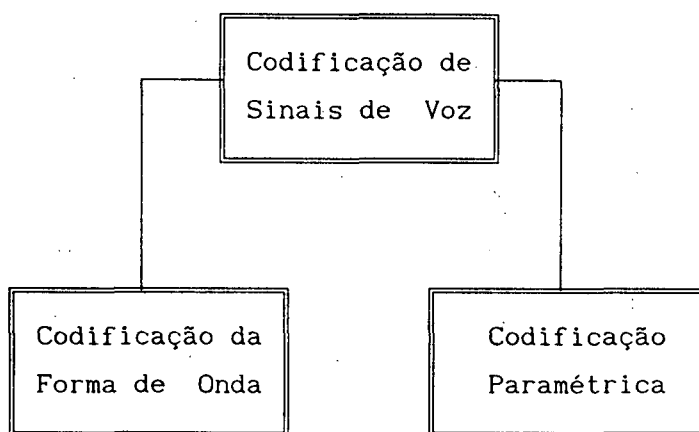


Fig.2.1 - Codificação de sinais de voz.

A codificação da forma de onda [11] consiste basicamente em um processo de aproximação da forma da onda do sinal de voz amostrado, segundo uma

determinada figura de mérito, como o erro instantâneo ou o erro médio quadrático, por exemplo. Já na codificação paramétrica, o sinal de voz é representado por parâmetros característicos, tais como composição espectral, pitch, energia, etc. Um dos procedimentos muito utilizados na codificação paramétrica, consiste em considerar que o sinal de voz pode ser representado pela resposta de um determinado filtro excitado por uma determinada entrada, restringindo-se então o processo de codificação à determinação dos parâmetros da fonte de excitação e de um filtro ou conjunto de filtros.

Um terceiro grupo de técnicas de codificação, ainda muito incipiente, é a codificação por regras, que pode também ser encarada como um processo de codificação paramétrica. A idéia deste tipo de codificação é a extração, na etapa de análise, de elementos fonéticos e de outros parâmetros tais como pitch e amplitude, que possam ser utilizados para representar convenientemente o sinal de voz. Os procedimentos de síntese por regras estão relativamente desenvolvidos, ao contrário dos procedimentos de análise, onde muito ainda deve ser investido. A expectativa para este tipo de codificação é a de serem obtidas taxas menores do que 100 bps para representação das locuções contidas em segmentos de voz [7,12].

Um aspecto relevante na escolha do procedimento de codificação de um sinal de voz é a taxa de transmissão ou de armazenamento. A Fig.2.2 apresenta uma distribuição de taxas em função do tipo de codificação para um sinal de voz com largura de banda igual a 3200 Hz [13]. A qualidade da voz sintetizada está expressa na escala MOS ("Mean Opinion Score") de avaliação subjetiva [13,14]. A meta atual das pesquisas (excluindo-se a codificação por regras), representada nesta figura, é a obtenção de codificadores que operem entre 4 kbps e 16 kbps, com qualidade subjetiva praticamente equivalente àquela obtida com o codificador log-PCM de 64 kbps.

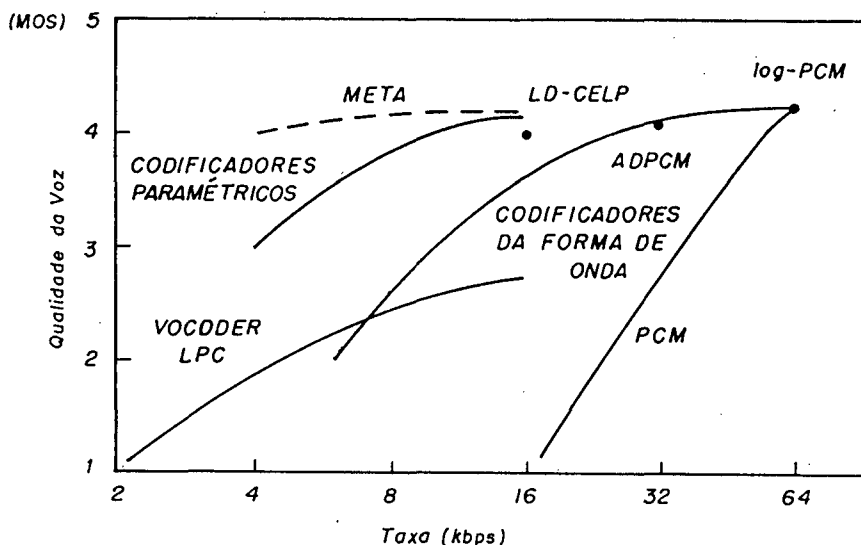


Fig.2.2 - Taxas de transmissão em função do tipo de codificação.

Além do tipo de codificação adotada, a taxa de transmissão e a qualidade do sinal sintetizado estão relacionadas com a limitação de banda no processo de amostragem. Em telefonia, a banda de voz é limitada em pouco mais de 3 kHz (300...3400 Hz), o que implica uma frequência de amostragem de 8 kHz. Para aplicações de alta fidelidade, taxas de amostragem da ordem de 20 kHz podem ser necessárias [1].

Neste capítulo, serão abordadas diversas técnicas para codificação de sinais de voz, sendo dada maior ênfase às técnicas de codificação paramétrica.

2.2 - CODIFICAÇÃO DA FORMA DE ONDA

Entre os procedimentos mais comuns de codificação da forma de onda de sinais de voz, podem ser citadas as codificações PCM, log-PCM, APCM, DPCM, ADPCM e a modulação delta.

A codificação PCM consiste basicamente na aproximação das amplitudes das amostras do sinal para um conjunto finito de valores, denominados níveis de

quantização, cada qual representado por uma sequência binária. Assim, um codificador PCM de B bits possui 2^B níveis distintos de amplitude uniformemente distribuídos para representar o sinal original, resultando em uma taxa de transmissão de $B.f_s$ bps, onde f_s é a frequência de amostragem em Hz, cujo valor mínimo é determinado pelo teorema da amostragem. Para obter uma representação de alta qualidade do sinal de voz na codificação PCM, são necessários em torno de 12 bits, o que para $f_s = 8$ kHz implica uma taxa de transmissão de 96 kbps.

Tendo em vista a disposição uniforme dos níveis de quantização, os segmentos de baixa amplitude na codificação PCM estão sujeitos a erros percentualmente mais significativos e portanto a uma relação sinal/ruído inferior. Uma alternativa para contornar este problema é o uso da compressão logarítmica do sinal de entrada, de forma a manter a relação sinal/ruído aproximadamente constante e independente do nível do sinal. A codificação log-PCM segundo as leis μ e A é comumente utilizada. As equações que representam as leis μ e A são encontradas em [11].

A codificação log-PCM com 8 bits utilizando as leis μ ou A apresenta aproximadamente a mesma qualidade da codificação PCM com 12 bits, com redução da taxa de transmissão de 96 kbps para 64 kbps, com $f_s = 8$ kHz [11].

Outros procedimentos de quantização não linear menos utilizados na prática consistem em comprimir o sinal de entrada do quantizador ou distribuir os níveis de quantização levando em conta a densidade de probabilidade deste sinal [11].

No PCM adaptável (APCM), a adaptação do quantizador segundo a variância do sinal de entrada permite a utilização de passos suficientemente largos para acomodar o nível máximo do sinal de entrada e de passos suficientemente pequenos de forma a minimizar o erro de quantização para os segmentos de baixo

nível. Tal objetivo pode ser alcançado com a aplicação de um controle de nível ao sinal de entrada de maneira que um sinal de variância aproximadamente constante seja aplicado a um quantizador de passo invariante no tempo. O uso de quantização adaptável implica uma melhora significativa na relação sinal/ruído em relação aos sistemas não adaptáveis.

Na codificação PCM diferencial (DPCM), é explorado o fato de que amostras adjacentes do sinal de voz são altamente correlacionadas, ou seja, o sinal varia pouco de amostra para amostra, sendo que a variância da diferença entre amostras adjacentes é menor do que a variância do sinal. Neste caso, o sinal a ser quantizado corresponde à diferença entre amostras do sinal original e de uma predição do mesmo. Sistemas DPCM adequadamente projetados implicam aumento de qualidade para uma mesma taxa em relação ao sistema PCM convencional, ou em menor taxa de transmissão sem degradação da qualidade.

Na codificação ADPCM [15], são combinadas as características dos codificadores APCM e DPCM, resultando em uma melhoria na relação sinal/ruído em relação ao log-PCM com mesmo número de bits. Podem ser utilizados os procedimentos "feedback" ou "feed-forward" para adaptação do passo de quantização e para determinação dos coeficientes do filtro encarregado da predição do sinal de entrada. No primeiro caso, os parâmetros são determinados a partir do próprio sinal sintetizado, dispensando a necessidade de transmissão dos mesmos ao decodificador. No procedimento "feed-forward", os parâmetros são determinados a partir do sinal de voz original e portanto devem ser enviados ao decodificador. Maiores ganhos na relação sinal/ruído podem ser obtidos na codificação ADPCM com a utilização de um preditor de longo atraso (PLA) que leve em consideração a correlação existente entre amostras separadas por um período de pitch [16,17].

O esquema geral para a modulação delta é semelhante ao do codificador

DPCM, sendo utilizada neste caso uma frequência de amostragem muito maior do que a frequência de Nyquist, resultando em amostras adjacentes altamente correlacionadas. O quantizador empregado possui apenas dois níveis (1 bit), o que implica uma taxa de transmissão igual à frequência de amostragem. No caso de ser utilizado um passo de quantização fixo, tem-se a chamada "Linear Delta Modulation" (LDM), que apresenta erros elevados para taxas de variação do sinal de entrada muito altas ou muito baixas. Um desempenho superior é obtido pelo codificador "Adaptive Delta Modulation" (ADM), onde o passo do quantizador é variante no tempo. Uma das vantagens da modulação delta é a simplicidade na implementação. Como desvantagem pode ser citada a necessidade de uma elevada taxa de amostragem para obtenção de um sinal decodificado de boa qualidade.

Na codificação da forma de onda podem ainda ser aplicados os procedimentos de quantização vetorial, sendo o sinal então dividido em blocos de amostras, cada qual representado por um índice que referencia uma sequência de um conjunto padrão ("codebook") [2].

2.3 - CODIFICAÇÃO PARAMÉTRICA

Os procedimentos de codificação paramétrica têm sido alvo de considerável quantidade de pesquisa, o que tem resultado no aumento da qualidade da voz sintetizada em taxas onde os codificadores da forma de onda não operam adequadamente, geralmente ao custo de maior complexidade.

A idéia básica da codificação paramétrica consiste em representar o sinal de voz através de determinados parâmetros, convenientemente escolhidos, que são utilizados pelo decodificador para reconstruir o sinal de voz.

Nesta seção é apresentado um modelo para síntese de voz resultante do

estudo das características do aparelho vocal humano, que é a base para uma grande classe de codificadores paramétricos. São apresentados ainda, de forma sucinta, alguns tipos de codificadores paramétricos, denominados vocoders ("voice" + "coders").

2.3.1 - MODELO BÁSICO PARA SÍNTESE DE VOZ

Uma das idéias básicas da codificação paramétrica, a representação dos sinais de voz como combinação de filtros e excitações variantes no tempo, está relacionada com a modelagem física do sistema vocal humano, assunto no qual destaca-se o trabalho de Gunnar Fant [18].

A Fig.2.3 apresenta um modelo físico aproximado para o sistema vocal humano [5]. O chamado trato vocal inicia na abertura das cordas vocais, denominada glote, e termina nos lábios, consistindo da faringe, que liga o esôfago à boca e da própria boca ou cavidade oral. O comprimento e a seção transversal do trato vocal variam de pessoa para pessoa e ainda de acordo com o som a ser emitido. O trato nasal começa no véu e termina nas narinas. Quando o véu está abaixado, o trato nasal é acusticamente acoplado ao trato vocal para produzir os sons nasais. Os pulmões, brônquios e traquéia constituem

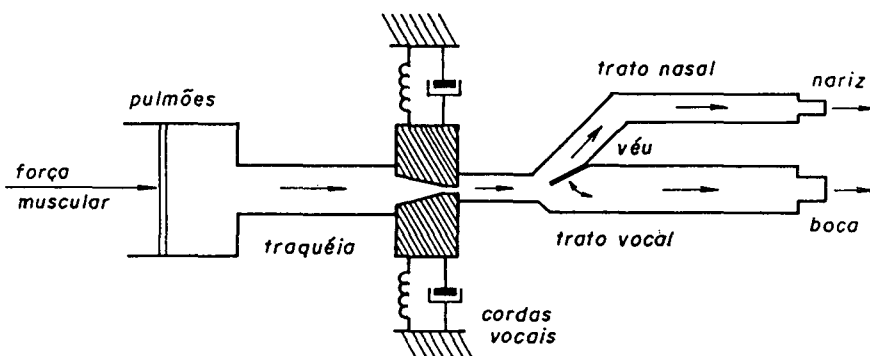


Fig.2.3 - Modelo físico aproximado do sistema vocal humano.

o sub-sistema glotal que serve de fonte de energia para produção da voz.

Os sons vocais podem ser classificados de acordo com o modo como são produzidos. Os sons vozeados são originados pela passagem de ar através da glote, fazendo as cordas vocais vibrarem produzindo pulsos de ar quase periódicos para a excitação do trato vocal. Quando o ar passa pelo orifício formado pelas cordas vocais, a pressão neste ponto cai fazendo com que a glote feche, aumentando a pressão sobre as cordas vocais até separá-las novamente. O período destes pulsos implica a característica da voz denominada pitch. Sons

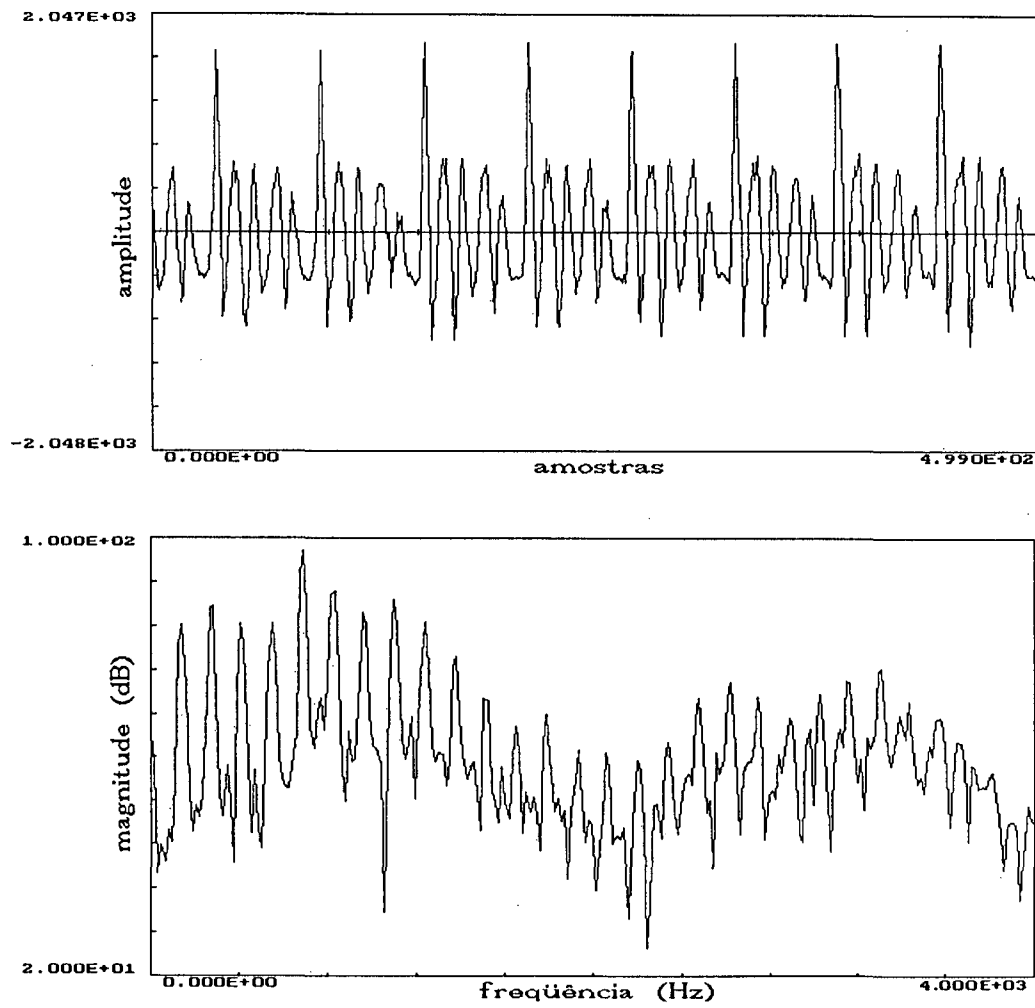


Fig.2.4 - Forma de onda e densidade espectral de potência da vogal "a".

fricativos são gerados pela passagem de ar através de uma constrição formada em algum ponto do trato vocal, com velocidade suficiente para produzir turbulência, tendo como efeito um ruído de espectro amplo. Os sons explosivos, também classificados como não vozeados, resultam de um aumento de pressão e do rápido alívio desta em algum ponto do trato vocal, determinando uma excitação do tipo impulsiva [1,6]. As Fig.2.4 e Fig.2.5 apresentam a forma de onda e a densidade espectral de potência para os sons correspondentes à vogal "a" (som vozeado) e à consoante "s" (som não vozeado), respectivamente, para uma frequência de amostragem igual a 8 kHz.

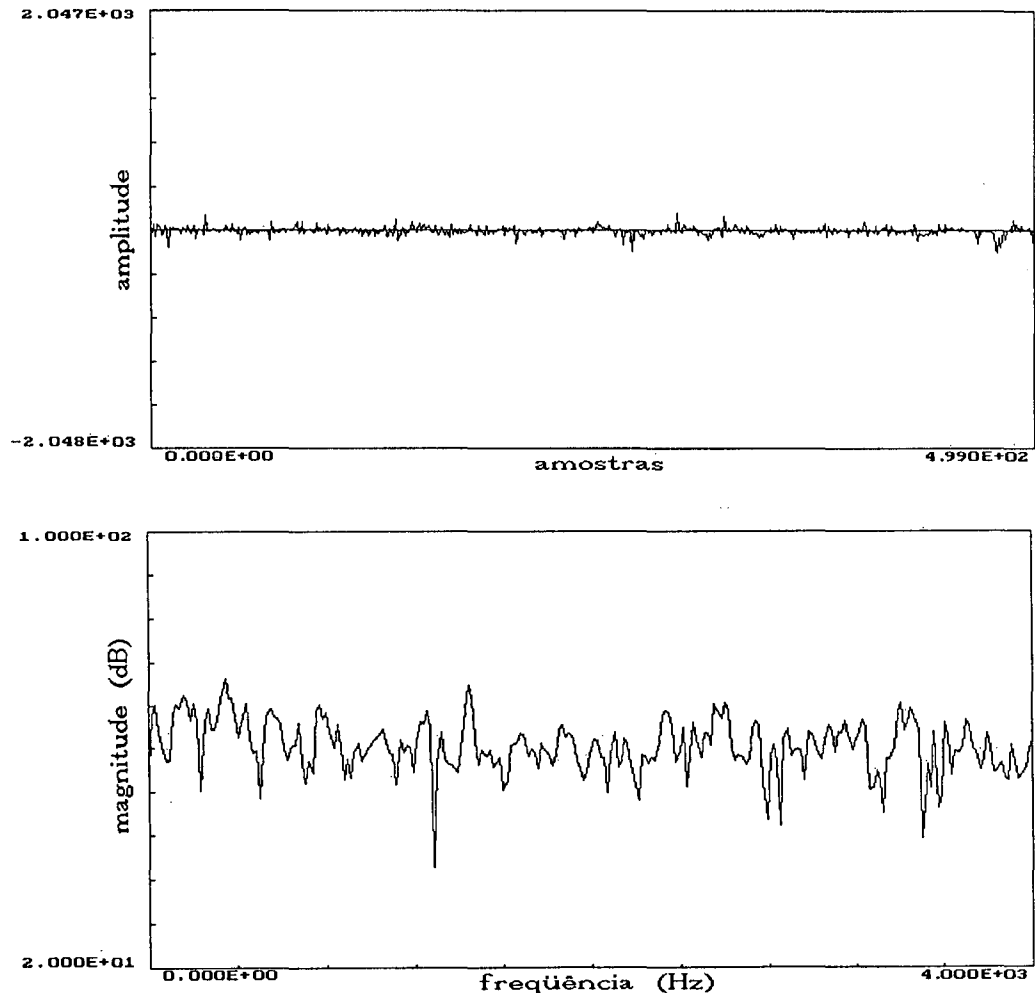


Fig.2.5 - Forma de onda e densidade espectral de potência da consoante "s".

O trato vocal e o trato nasal podem ser vistos como tubos com áreas transversais não uniformes. A variação da forma destes tubos implica diferentes frequências de ressonância que determinam o envelope espectral do som produzido. As frequências de ressonância do tubo formado pelo trato vocal são chamadas de formantes, e sua largura de banda é influenciada pelas perdas térmicas, por fricção viscosa e pelas perdas de radiação [1]. A dinâmica das dimensões do trato vocal implica que o sinal de voz possua uma característica espectral variante no tempo.

Nos sons nasais, o véu é totalmente abaixado e o fluxo de ar decorrente da excitação glotal é irradiado pelas narinas. A cavidade oral funciona como uma cavidade ressonante produzindo anti-ressonâncias ou zeros no som gerado. A Fig.2.6 apresenta um modelo para esta situação.

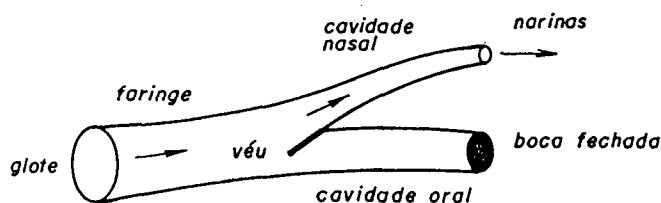


Fig.2.6 - Modelo de tubos para produção de sons nasais.

O modelo físico para o aparelho vocal humano apresentado é aproximado. A elaboração de um modelo completo para o desenvolvimento da teoria acústica da voz, o que não é tarefa simples, deve considerar os seguintes efeitos [1]:

1. variação temporal da forma do trato vocal;
2. perdas devidas à condução de calor e fricção viscosa nas paredes do trato vocal;
3. suavidade das paredes do trato vocal;
4. radiação do som dos lábios;

5. acoplamento nasal;
6. excitação do som no trato vocal.

Em uma forma mais simplificada, o sistema vocal pode ser visto como um sistema linear e variante no tempo, cuja resposta, quando adequadamente excitado, corresponde ao sinal de voz. A Fig.2.7 apresenta o referido modelo. Uma simplificação adotada é a independência entre a excitação e o sistema, o que não chega a comprometer a aceitação deste modelo como base para o desenvolvimento de uma importante classe de codificadores paramétricos.

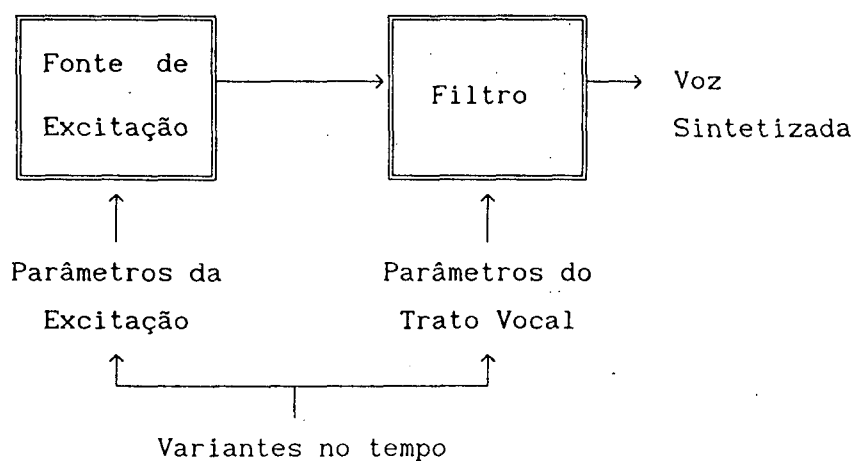


Fig.2.7 - Modelo fonte/filtro para produção de voz.

Para muitos sons de voz é razoável assumir que as propriedades estatísticas da excitação e do trato vocal permanecem constantes para intervalos da ordem de 10 a 30 ms [10], o que implica a necessidade de atualização dos parâmetros do modelo em intervalos desta ordem.

Com a popularização das técnicas digitais, a simulação de modelos como o apresentado anteriormente fica grandemente facilitada. Neste caso, um sistema discreto gera uma seqüência de amostras com características básicas de um

sinal de voz amostrado.

Considerando os tipos de excitação que produzem os sons de voz e os diversos efeitos que dão forma ao sinal de voz, pode ser imaginado um modelo mais completo como o apresentado na Fig.2.8 [1].

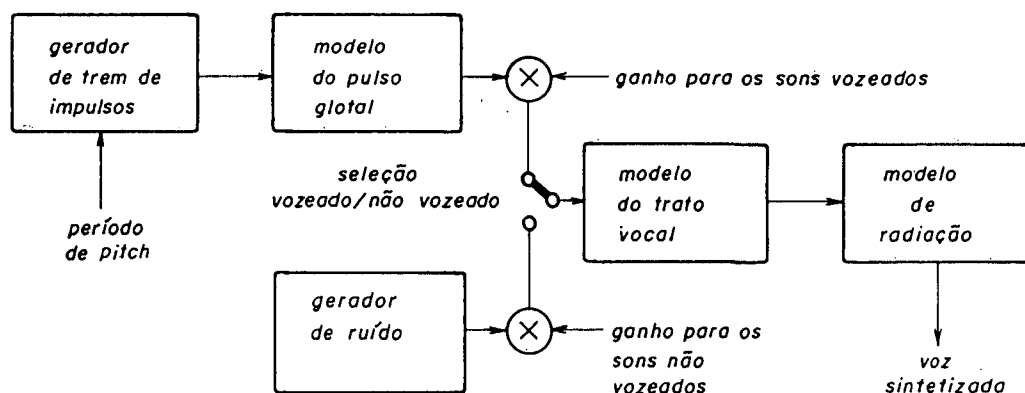


Fig.2.8 - Modelo para síntese de voz.

O trato vocal pode ser modelado por um filtro auto-regressivo do tipo "all-pole", cujos coeficientes determinam as frequências e larguras de banda dos formantes. O referido filtro pode ser representado pela função de transferência $F(z)$ dada a seguir:

$$F(z) = \frac{1}{1 - \sum_{i=1}^{OP} a_i \cdot z^{-i}} \quad (2.1)$$

onde:

OP é a ordem do filtro;

a_i são os coeficientes do filtro.

As ressonâncias do trato vocal (formantes) estão diretamente relacionados com os pólos da função de transferência $F(z)$. O modelamento de

sons nasais e fricativos requer ressonâncias e anti-ressonâncias (pólos e zeros). Neste caso, podem ser incluídos zeros ou mais pólos em $F(z)$ para simular o efeito dos zeros [19].

O modelo de radiação considera o efeito da radiação acústica nos lábios. Uma aproximação razoável consiste em utilizar um filtro ARMA ("Auto Regressive-Moving Average") que pode ter seu efeito embutido em $F(z)$ da mesma forma que as anti-ressonâncias do trato vocal [1].

Tendo em vista que os sons de voz podem ser divididos de uma maneira geral em vozeados e não vozeados, a excitação deve ser capaz de fornecer uma seqüência quase periódica de pulsos ou um ruído branco, conforme o caso. O trem de impulsos (ou amostras unitárias) com o período de pitch é convolvido com uma função que modela a forma do pulso glotal, gerando uma excitação do tipo representado na Fig.2.9.

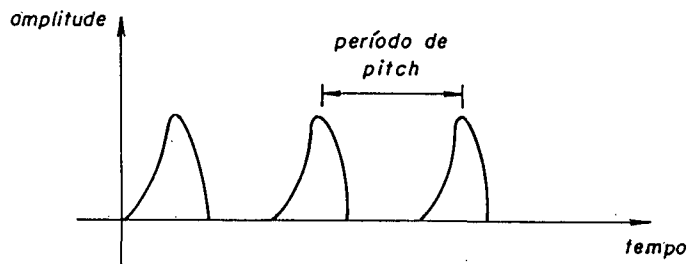


Fig.2.9 - Forma aproximada da excitação glotal.

A função de transferência que modela o pulso glotal pode ser bem representada por uma função com somente zeros finitos, entretanto, um modelo de dois pólos tem sido utilizado com sucesso [1].

Nos sons não vozeados, a excitação é um ruído branco. Para os modelos discretos, um gerador de números pseudo-aleatórios fornece uma fonte de ruído com envelope espectral praticamente plano.

A amplitude do sinal de voz gerado é controlada por dois ganhos que ponderam as componentes vozeadas e não vozeadas da excitação.

Em alguns casos, como na análise preditiva linear (LPC), é conveniente combinar os efeitos do trato vocal, do pulso glotal e do modelo de radiação em um único modelo do tipo só pólos. Este objetivo pode ser alcançado através do aumento do número de pólos de $F(z)$.

O modelo de produção de voz apresentado possui algumas limitações. Primeiro, os sons transientes não são adequadamente produzidos tendo em vista a invariância temporal dos parâmetros durante intervalos de tempo relativamente longos. Segundo, a falta de zeros na modelagem das funções limita a qualidade do sons nasais e fricativos. Terceiro, a opção radical entre sons vozeados e não vozeados não é adequada para alguns tipos de sons de voz. Por último, o modelo requer que os pulsos glotais sejam espaçados por um múltiplo inteiro do período de amostragem [1].

2.3.2 - VOCODERS

Um importante grupo de codificadores de voz são os chamados vocoders, onde o sinal de voz é codificado e reconstituído a partir de um grupo de parâmetros variantes no tempo. O principal objetivo dos vocoders é o de obter uma representação discreta do sinal de voz a taxas muito menores do que aquelas obtidas na codificação da forma de onda. Sua flexibilidade permite alteração dos parâmetros do sinal de voz mais facilmente do que com outras formas de codificação. Dentro deste conjunto de codificadores encontram-se o vocoder de canal, o vocoder de fase, o vocoder homomórfico, o vocoder de formantes, o vocoder LPC e toda a classe de codificadores de análise-por-síntese, onde inclui-se o codificador CELP.

O mais antigo dispositivo de codificação de voz é o chamado vocoder de canal de Dudley. A Fig.2.10 apresenta o diagrama em blocos da etapa de análise do vocoder de canal. Cada canal possui um filtro passa-banda centrado em uma frequência diferente, de forma a cobrir toda a faixa do sinal de entrada. O extrator de magnitude e o filtro passa-baixas funcionam como um detector do envelope da magnitude da transformada de Fourier na frequência do canal. O dizimador está presente pelo fato de que a saída de cada canal pode ser amostrada a uma taxa inferior àquela utilizada para amostrar o sinal de entrada. O analisador inclui ainda um detector de sons vozeados e um detector de pitch. No processo de síntese, os sinais de magnitude controlam as amplitudes das excitações (ruído ou trem de impulsos) de uma série de filtros passa-banda, cuja soma das saídas fornece o sinal de voz sintetizado. O tipo da excitação é definida pelas informações vozeado/não vozeado e pelo período de pitch. A voz resultante deste processo é altamente reverberante [1].

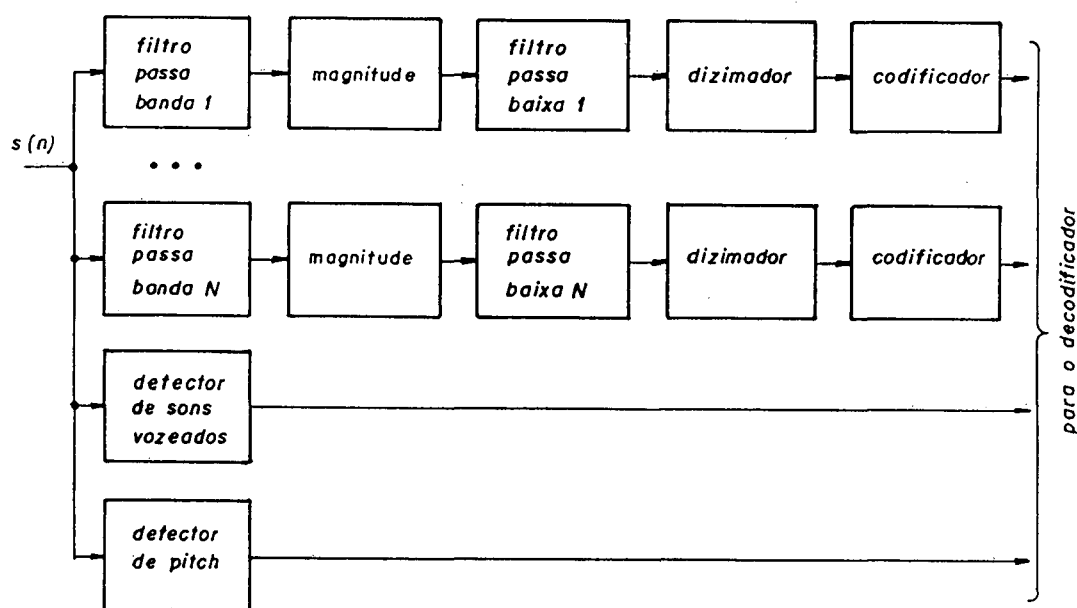


Fig.2.10 - Vocoder de canal (codificador).

Os vocoders de canal operam tipicamente na faixa entre 1200 e 9600 bps, com aproximadamente 600 bps dedicados às informações de som vozeado/não vozeado e do período de pitch. A manipulação direta do pitch implica uma vantagem quanto à flexibilidade e uma desvantagem quanto à complexidade, já que a determinação precisa do mesmo é freqüentemente uma tarefa difícil [1].

A Fig.2.11 apresenta o diagrama do processo de análise de um canal do vocoder de fase. Cada canal é centrado em uma freqüência ω_k . O efeito do recobrimento na amostragem da magnitude e da derivada da fase são menos percebidos do que o efeito da amostragem das partes real e imaginária da transformada de Fourier do sinal de entrada. A presença dos dizimadores indica que tanto o módulo como a derivada da fase podem ser amostrados com uma freqüência inferior àquela utilizada para o sinal de entrada.

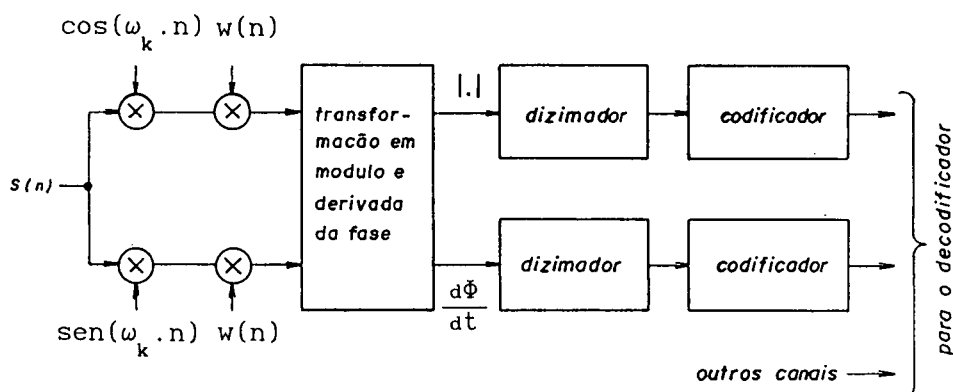


Fig.2.11 - Canal de um vocoder de fase (codificador).

No vocoder de fase, a magnitude do sinal de saída do analisador está relacionada com a transferência do trato vocal, enquanto que a fase ou derivada da fase está relacionada com a excitação. O sinal sintetizado é obtido no decodificador como a soma das saídas de vários canais [1].

No vocoder homomórfico, o pitch e a detecção dos sons vozeados é realizada com base no cepstro do sinal calculado a cada intervalo de 10 a 20 ms. O cepstro é definido como a transformada de Fourier inversa do logaritmo natural da transformada de Fourier do sinal [1]. Neste caso, a idéia básica do processamento homomórfico é a de separar dois sinais, a excitação e a resposta ao impulso, combinados através de convolução. Os baixos tempos dos cepstros são quantizados e codificados para transmissão ou armazenagem. No sintetizador, os baixos tempos do cepstro são utilizados para gerar uma aproximação da resposta ao impulso que combina os efeitos do trato vocal, pulso glotal e efeitos de radiação, que será convolvida com uma excitação adequada para gerar o sinal de voz. A Fig.2.12 apresenta o diagrama em blocos do codificador do vocoder homomórfico.

O vocoder homomórfico é declarado como sendo capaz de produzir voz de muito alta qualidade com 26 amostras do cepstro quantizadas com 6 bits e atualizadas 50 vezes por segundo [1].

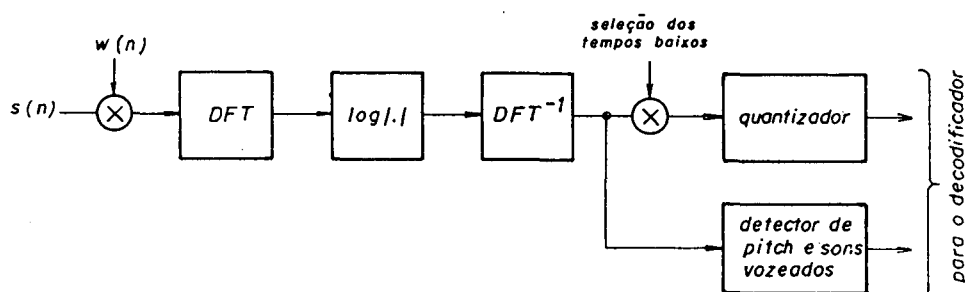


Fig.2.12 - Vocoder homomórfico (codificador).

O diagrama de blocos do vocoder de formantes é apresentado na Fig.2.13. Neste modelo, os sons vozeados são caracterizados pelo período de pitch, pela amplitude e pelos três primeiros formantes. Os sons não vozeados são

caracterizados pela amplitude e por um sistema zero/pólo. Uma compensação fixa aproxima o pulso glotal e os efeitos de radiação. Se o som é vozeado, o que é decidido pela análise do cepstro, o período de pitch é estimado também com base no cepstro e os três primeiros formantes são estimados com base nos picos do logaritmo do espectro. No caso de sons não vozeados, o pólo é localizado no maior pico do logaritmo do espectro e o zero é localizado de maneira a manter a amplitude relativa entre altas e baixas frequências. Os parâmetros são estimados 100 vezes por segundo e filtrados. As larguras de banda dos formantes são mantidas fixas. Os parâmetros podem ser transmitidos com uma largura de banda de aproximadamente 16 Hz sem perda apreciável de qualidade. Utilizando-se uma taxa de amostragem dos mesmos em torno de 33 amostras/segundo, obtém-se uma taxa em torno de 700 bps [1].

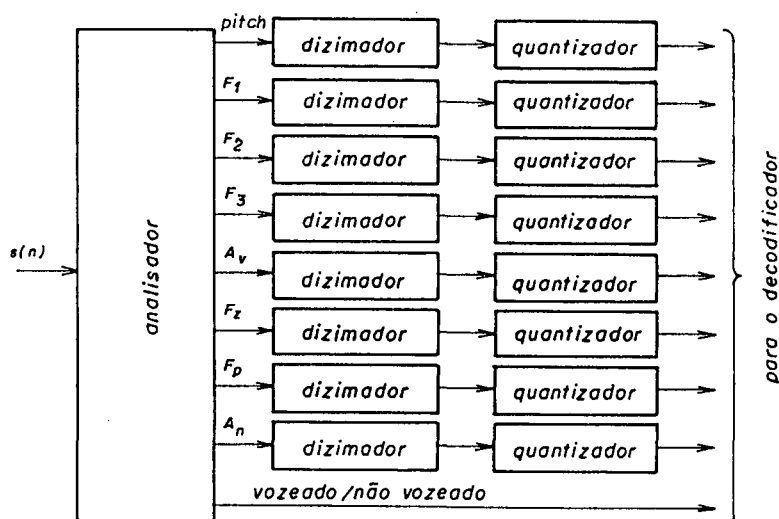


Fig.2.13 - Vocoder de formantes (codificador).

Uma das mais importantes aplicações da análise preditiva linear é na área da codificação de voz [20]. A Fig.2.14 mostra o diagrama em blocos do codificador de um vocoder LPC. O transmissor realiza uma análise LPC para

determinação do envelope espectral do sinal de voz. Os parâmetros transmitidos são um conjunto de OP coeficientes de predição, o período de pitch, identificação de som vozeado/não vozeado e o ganho. A quantização direta dos coeficientes do preditor pode levar a grandes mudanças nas posições dos pólos e inclusive à instabilidade do filtro. Como alternativas podem ser transmitidas as raízes do polinômio do preditor, os coeficientes de reflexão ou os coeficientes PARCOR [1].

A utilização de representação PCM para os parâmetros do preditor conduz a taxas tipicamente entre 2400 e 7200 bps, enquanto que a representação ADPCM pode conduzir a uma voz de boa qualidade com taxas entre 1000 e 2000 bps [1].

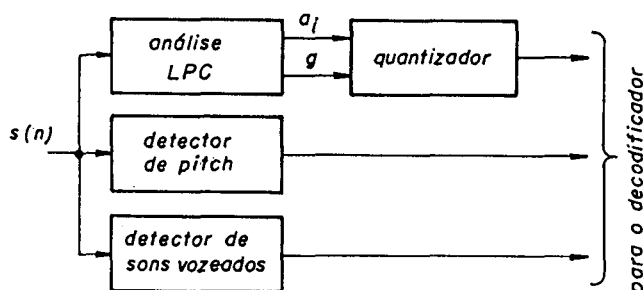


Fig.2.14 - Vocoder LPC (codificador).

Um dos principais problemas em alguns vocoders é a estimação da função excitação. Uma aproximação para contornar a determinação direta dos parâmetros da excitação é o vocoder LPC excitado por voz apresentado na Fig.2.15. Neste caso, as componentes de baixa frequência do sinal de voz são codificadas por uma técnica de codificação de forma de onda e utilizadas no sintetizador para gerar a excitação. A justificativa para este procedimento é a de que o sinal de baixa frequência contém toda a informação necessária sobre a excitação, ou seja, tem forma periódica para os sons vozeados e forma de ruído para os sons

não vozeados. Além dos parâmetros da excitação, também são transmitidos os parâmetros do preditor LPC obtidos do sinal de entrada. A taxa requerida pelo vocoder LPC excitado por voz é cerca de 1000 a 2000 bps maior do que a requerida pelo vocoder LPC, porém a qualidade da voz obtida é melhorada devido à eliminação da estimação do pitch e da detecção de sons vozeados [1]. Uma forma de excitação mista para o vocoder LPC é apresentada em [21].

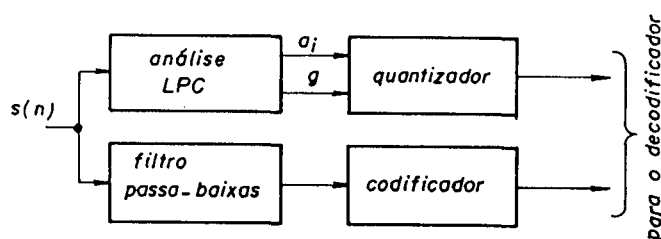


Fig.2.15 - Vocoder LPC excitado por voz (codificador).

Nos codificadores que utilizam a técnica de análise por síntese [22], a excitação é determinada de forma a minimizar a energia da diferença entre o segmento de voz sintetizado e o segmento de voz original. Dentre as inúmeras estruturas incluídas nesta classe, podem ser destacadas algumas básicas, como o codificador CELP ("Code-Excited Linear Predictive"), o MPELPC ("Multi-Pulse Excited Linear Predictive Coding") e o SEV ("Self-Excited Vocoder"). Todos os codificadores mencionados diferem basicamente quanto à excitação, sendo em todos os casos, o envelope espectral do sinal de voz obtido através de um filtro determinado por técnicas de predição linear.

No codificador CELP convencional, a melhor excitação é escolhida através de uma busca exaustiva em um conjunto de seqüências pseudo-aleatórias, sendo transmitido somente o índice referente à excitação escolhida. A periodicidade na excitação é obtida através de um filtro, denominado filtro preditor de longo atraso, cujos parâmetros podem também ser determinados de forma a

minimizar a energia do erro entre o sinal sintetizado e o sinal original.

No codificador MPELPC [23-7], a excitação é formada por algumas poucas amostras, cujas amplitudes e posições dentro do segmento de excitação são convenientemente determinadas. Neste tipo de codificador, a correta alocação dos pulsos pode prover a periodicidade adequada na excitação, entretanto, a utilização de um filtro PLA conduz a melhorias na qualidade do sinal sintetizado [25].

No codificador SEV, a excitação é formada através de uma cascata de dois filtros PLA ou de um único filtro PLA. A saída atual de cada filtro PLA é função unicamente de seus valores passados (resposta à entrada zero), não existindo excitação explícita [22,28,29].

2.4 - CONCLUSÕES

As técnicas de codificação de voz podem ser classificadas em dois grandes grupos:

- codificação da forma de onda;
- codificação paramétrica.

Dentro desta classificação, as técnicas de codificação paramétrica têm sido alvo de grande quantidade de pesquisa, dada a possibilidade de obtenção de voz de boa qualidade em taxas mais baixas do que aquelas obtidas com os codificadores de forma de onda.

As técnicas de codificação da forma de onda têm em seu favor a menor complexidade computacional, sobretudo na etapa de codificação, o que facilita a implementação de codificadores/decodificadores que operem em tempo real, fornecendo um sinal sintetizado de alta qualidade.

Além das técnicas de codificação de sinais de voz consagradas e

incluídas nestes dois grupos, as técnicas de codificação por regras, ainda incipientes, principalmente com relação ao processo de análise, apresentam-se como uma possibilidade de codificação de sinais de voz a taxas muito baixas.

Pode ser afirmado que, de uma forma geral, a complexidade dos codificadores cresce com a diminuição da taxa de transmissão desejada. Neste ponto, ressalta-se que o problema da complexidade computacional tem sido amenizado pela disponibilidade de processadores de sinais cada vez mais rápidos e econômicos.

Com relação aos diversos tipos de vocoders apresentados, a comparação dos desempenhos quanto à qualidade do sinal sintetizado, taxa e complexidade, fica prejudicada, já que as referências consultadas não são conclusivas quanto a este respeito. O estudo mais aprofundado destas estruturas é apresentado como uma sugestão para futuros trabalhos.

Dentre os diversos codificadores de voz mencionados, o codificador CELP, incluído na categoria dos codificadores paramétricos que utilizam o procedimento de análise-por-síntese, é um modelo de grande sucesso, dada a boa qualidade de voz obtida em taxas tão baixas quanto 4800 bps [30]. Sua principal aplicação, os sistemas de comunicação telefônica, exige possibilidade de operação em tempo real, o que torna a redução da complexidade computacional deste modelo um objetivo bastante perseguido, mesmo diante do desempenho apresentado pelos processadores de sinais disponíveis atualmente [31-33].

CAPÍTULO 3

CODIFICAÇÃO CELP3.1 - INTRODUÇÃO

A idéia básica do codificador CELP convencional ("Code-Excited Linear Predictive") foi inicialmente proposta por Atal e Schroeder em 1982 [34,35]. Sua característica principal consiste na quantização vetorial da excitação através da procura entre um conjunto de excitações aleatórias, daquela que resulte na menor energia da diferença entre o sinal sintetizado e o sinal de voz original. A Fig.3.1 apresenta o diagrama em blocos da etapa de análise do codificador CELP convencional. A etapa de síntese é apresentada na Fig.3.2.

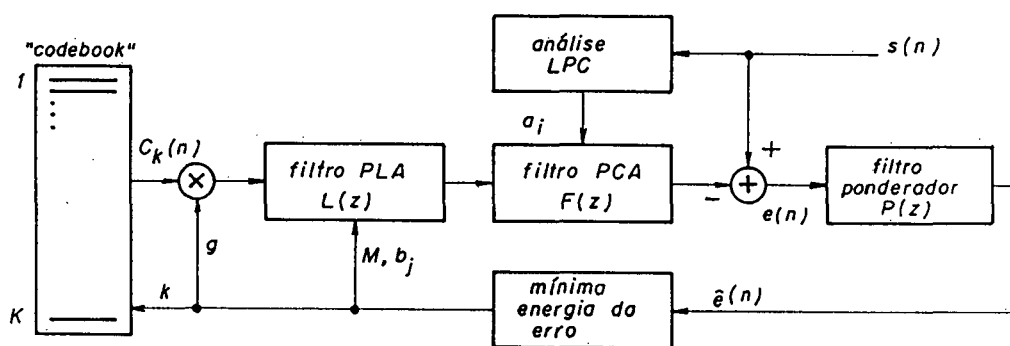


Fig.3.1 - Modelo de análise do codificador CELP convencional.

O filtro preditor de curto atraso (PCA) é responsável pela reconstituição das correlações entre amostras pouco distanciadas no sinal de

voz original. A determinação de seus coeficientes é normalmente realizada através de técnicas LPC ("Linear Predictive Coding"), em intervalos denominados quadros.

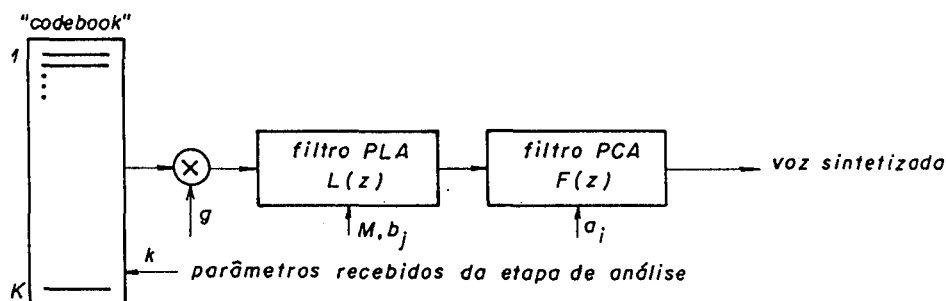


Fig.3.2 - Modelo de síntese do codificador CELP convencional.

O filtro preditor de longo atraso (PLA) é responsável pela introdução de uma periodicidade no sinal de excitação do filtro PCA, relacionada com o pitch. Seus parâmetros, um atraso M e um ou mais coeficientes, podem ser determinados de forma a minimizar a energia da diferença ponderada, $\hat{e}(n)$, entre o sinal de voz original e o sinal sintetizado, em intervalos denominados subquadros.

O "codebook" é composto normalmente por um conjunto de K seqüências de duração igual à da excitação, com distribuição gaussiana e variância unitária [34]. A determinação da seqüência ("codeword") que melhor represente a excitação é feita de forma a minimizar a energia de $\hat{e}(n)$ dentro de um subquadro. A utilização de "codebooks" não estocásticos também tem sido adotada, tal como em [36], onde o "codebook" é projetado através do algoritmo LBG (Linde, Buzo e Gray) [37]. Em [38] é proposto um modelo que utiliza um "codevector" não estocástico, caso em que o índice indica a posição de uma seqüência de amostras dentro do vetor de excitações.

O ganho g faz o ajuste da amplitude das amostras da "codeword" reduzindo a energia de $\hat{e}(n)$ e aumentando virtualmente o tamanho do "codebook", sem contudo aumentar proporcionalmente a complexidade na determinação da melhor "codeword".

No modelo de síntese, os parâmetros referentes aos coeficientes dos filtros PCA e PLA, o atraso M do filtro PLA, o ganho e o índice da "codeword" são combinados de maneira a obter-se o sinal sintetizado.

O filtro ponderador do erro é responsável pela concentração do ruído nas regiões de maior energia do envelope espectral (regiões dos formantes), onde ele é subjetivamente menos perturbador, pois é mascarado pelos formantes, aumentando assim a qualidade subjetiva do sinal sintetizado.

Existem várias estruturas de codificação baseadas no codificador CELP convencional com o objetivo de obter um melhor desempenho, principalmente com relação à complexidade computacional.

O VXC ("Vector Excitation Coding") de Allen Gersho [39] tem a mesma estrutura que o codificador CELP convencional.

Em [40,41] é apresentado o ACELP ("Adaptive Codebook CELP"), onde é utilizado um "codebook" que é adaptado com base nos segmentos de voz passados e que tem como objetivo a eliminação da filtragem de cada candidata na busca da melhor "codeword", diminuindo assim a complexidade.

O BACELP ("Backward block Adapted CELP") [42], o LD-CELP ("Low Delay - CELP") [43] e o LD-VXC ("Low Delay - VXC") [44] são codificadores de baixo atraso que utilizam adaptação dos parâmetros do tipo retroativa ("backward") e que por este motivo devem operar em taxas mais altas (em torno de 16 kbps). Os únicos parâmetros enviados ao decodificador são o índice da "codeword" e o ganho [45], ou somente o índice da "codeword" [46,47].

O SPE-CELP ("Single-Pulse Excitation CELP") utiliza uma estrutura

semelhante à do CELP para os sons não vozeados. Para os sons vozeados é utilizada uma excitação baseada em pulsos espaçados por períodos de pitch. Esta estrutura tem desempenho superior ao do CELP convencional em taxas menores do que 4 kbps [48].

O Regular-Pulse Excitation [49,50] é um codificador que apresenta a excitação composta por amostras unitárias regularmente espaçadas e um vetor de ganhos associado. Seu princípio assemelha-se mais ao do codificador Multi-Pulso. Idéia basicamente semelhante é utilizada em [51] onde foi verificado um desempenho superior ao do CELP convencional em 4800 bps.

A densidade dos pulsos de excitação constante dentro de um subquadro, mas variando de subquadro para subquadro, é a principal característica do ADP-CELP ("Adaptive Density Pulse CELP"), sendo sua complexidade inferior à do CELP convencional, tendo também apresentado melhores resultados objetivos [52,53].

O CVXC ("Classified Vector Excitation Coding") utiliza vários "codebooks", dos quais somente um é escolhido com base nas características do resíduo LPC. Esta estrutura apresentou menor complexidade do que o CELP convencional, tendo sido também verificada a melhor reprodução de picos na excitação [54].

A idéia de utilizar vários "codebooks" de tamanho reduzido para formar a excitação é utilizada pelo VSELP ("Vector Sum Excited Linear Prediction") [2,55] e pelo MSVXC ("Multiple-Stage Vector Excitation Coding") [56] com o objetivo de reduzir a complexidade computacional com manutenção da qualidade final [56].

Em [57] é utilizada uma estrutura cujo princípio é semelhante ao do codificador CELP, tendo sido obtidos bons resultados na codificação de áudio de

alta qualidade (largura de banda em torno de 15 kHz) com taxas menores do que 100 kbit/s.

A divisão do sinal de voz em sub-bandas e a aplicação da codificação CELP a cada uma delas é proposta em [58]. Idéia semelhante é utilizada em [59] para sinais de voz com largura de banda de 7500 Hz e amostrados a 16 kHz.

É importante ressaltar que algumas referências não são conclusivas a respeito das vantagens e limitações de suas propostas em relação ao codificador CELP convencional. O estudo mais aprofundado de cada uma das estruturas mencionadas apresenta-se como um tema interessante para trabalhos futuros.

3.2 - DETERMINAÇÃO DOS PARÂMETROS DO CODIFICADOR CELP CONVENCIONAL

A seguir são apresentadas várias técnicas para determinação dos parâmetros do codificador CELP convencional. São descritos também os procedimentos de ponderação do erro, expansão da banda do filtro PCA e pós-filtragem, os quais têm como objetivo a melhoria da qualidade do sinal de voz sintetizado.

3.2.1 - DETERMINAÇÃO DOS COEFICIENTES DO FILTRO PCA

Via de regra, é adotado um modelo auto-regressivo (AR) para o filtro PCA, cujos coeficientes podem ser determinados através de técnicas LPC [60]. Em [61] é adotado um filtro ARMA ("AutoRegressive Moving Average"), enquanto que em [62], técnicas homomórficas são utilizadas na determinação dos coeficientes do filtro PCA. A otimização dos coeficientes do filtro PCA após a determinação da excitação é proposta em [63].

A utilização de um preditor bilateral simétrico, onde as amostras são preditas com base em valores de amostras passadas e futuras dentro do quadro de análise, não conduziu a resultados satisfatórios em nossas simulações com o codificador CELP. A vantagem da utilização de um preditor bilateral simétrico prende-se ao fato de ser necessária a transmissão de somente a metade dos coeficientes [64].

A utilização de um filtro PCA do tipo ARMA, com determinação simultânea dos coeficientes do numerador e denominador pelo método proposto em [65,66] também não apresentou resultados satisfatórios em nossas simulações com o codificador CELP.

Considerando a modelagem do filtro PCA como um filtro AR, a resposta do filtro PCA AR inverso, quando excitado pelo sinal de voz $s(n)$, é dada por:

$$r(n) = s(n) - \sum_{i=1}^{OP} a_i \cdot s(n-i) \quad (3.1)$$

onde:

OP é a ordem do filtro PCA.

Os coeficientes a_i são determinados de forma a minimizar a energia E de $r(n)$ dentro de um determinado intervalo de amostras. Assim:

$$E = \sum_n r^2(n) = \sum_n \left[s(n) - \sum_{i=1}^{OP} a_i \cdot s(n-i) \right]^2 \quad (3.2)$$

A minimização da energia de $r(n)$ ocorre quando as derivadas parciais de E em relação aos coeficientes a_i são igualadas a zero. Tal condição conduz ao conjunto de equações lineares, conhecido como equações normais [60], apresentado a seguir:

$$\sum_{i=1}^{OP} a_i \cdot \sum_n [s(n-i) \cdot s(n-k)] = \sum_n [s(n) \cdot s(n-k)] \quad , 1 \leq k \leq OP \quad (3.3)$$

Dependendo dos limites em n impostos para os somatórios em (3.3), podem ser aplicados os métodos da autocorrelação e da covariância, apresentados a seguir.

a) MÉTODO DA AUTOCORRELAÇÃO

Considerando que a energia E do erro $r(n)$ seja minimizada no intervalo $0 < n < N_q - 1 + OP$, onde N_q é a duração do quadro de análise, o conjunto de equações normais é reduzido a:

$$\sum_{i=1}^{OP} a_i \cdot R(|i-k|) = R(k) \quad , 1 \leq k \leq OP \quad (3.4)$$

onde:

$R(\tau)$ é a autocorrelação do sinal de voz $s(n)$ em τ .

O conjunto de equações representado por (3.4) é conhecido como equações de Yule-Walker [60].

A matriz formada do lado esquerdo de (3.4) é simétrica e Toeplitz (todos elementos iguais ao longo de cada diagonal) e portanto gera um sistema que pode ser solucionado pelo método de Durbin [1]. A utilização deste procedimento pode conduzir a sistemas mal condicionados [60], o que não chega felizmente a ser um inconveniente na prática. Em nossas simulações foi possível observar tal fenômeno, visto terem sido obtidos resultados ligeiramente diferentes ao compilar e executar o programa em diferentes máquinas, ou em uma mesma máquina, utilizando diferentes opções de compilação (alteração de precedência nas operações algébricas).

Além do ganho de velocidade obtido pelo emprego do método de Durbin, a utilização de formas rápidas na determinação das autocorrelações pode aumentar a eficiência do método da autocorrelação. No Apêndice A é apresentado um procedimento rápido sistematizado para o cálculo de autocorrelações, proposto por Blankinship [67]. Um método para o cálculo das autocorrelações que implica redução do número de multiplicações com aumento do número de adições é encontrado em [68].

Outro aspecto relevante do método da autocorrelação diz respeito à garantia de estabilidade do filtro PCA.

A atualização dos coeficientes do filtro PCA é realizada em segmentos (quadros) da ordem de 20 a 30 ms, aproveitando a característica quase estacionária do sinal de voz em intervalos desta ordem. Para melhorar a estimativa das autocorrelações em vista da segmentação do sinal, é usual tomar um intervalo de amostras maior do que um quadro e ponderar o mesmo de forma que as amostras situadas nas extremidades tenham menor influência. Este procedimento é ilustrado na Fig.3.3.

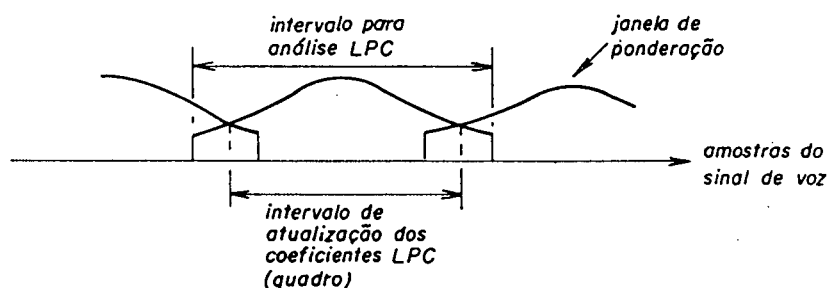


Fig.3.3 - Procedimento de segmentação e ponderação do sinal de voz para análise LPC.

Via de regra é adotada a janela de Hamming para ponderação das amostras

do sinal de voz na análise LPC quando utilizado o método da autocorrelação. O janelamento recursivo de Barnwell [69] foi utilizado em [46] para determinação das autocorrelações, com redução do espaço de memória necessário. Uma janela do tipo trapezoidal é utilizada em [70].

b) MÉTODO DA COVARIÂNCIA

Caso a minimização do erro ocorra no intervalo $0 \leq n \leq N_q$, as equações normais são reduzidas a:

$$\sum_{i=1}^{OP} a_i \cdot \Phi_{ss}(i,k) = \Phi_{ss}(0,k) \quad , 1 \leq k \leq OP \quad (3.5)$$

onde

$$\Phi_{ss}(i,k) = \sum_{n=0}^{N_q-1} [s(n-i) \cdot s(n-k)] \text{ é a covariância de } s(n); \quad (3.6)$$

A matriz obtida do lado esquerdo de (3.5) é simétrica mas não Toeplitz [60], já que:

$$\Phi_{ss}(i+1,k+1) = \Phi_{ss}(i,k) + s(-i-1) \cdot s(-k-1) - s(N_q-1-i) \cdot s(N_q-1-k) \quad (3.7)$$

Neste caso, um método eficiente para resolução do sistema é a decomposição de Cholesky [1], sistematizada no Apêndice B. A propriedade representada por (3.7) pode ser aplicada na montagem do sistema de equações, com o objetivo de reduzir a complexidade do procedimento.

Uma desvantagem deste método é a não garantia de estabilidade do filtro PCA. A estabilização pode ser realizada através da transformação dos coeficientes a_i em coeficientes PARCOR, com a utilização do método de Durbin,

por exemplo, e limitação dos mesmos entre -1 e 1 , que é condição suficiente para estabilidade do filtro [60]. Os PARCOR limitados são então transformados em coeficientes do filtro PCA estabilizado através do método de Durbin. O método da covariância modificado desenvolvido por Atal [71,72] é baseado no procedimento descrito.

No método da covariância não existe necessidade de ponderação do sinal de voz com uma janela não retangular, sendo porém necessário o conhecimento das amostras do sinal $s(n)$ no intervalo $-OP \leq n \leq Nq-1$.

3.2.2 - DETERMINAÇÃO DOS PARÂMETROS DO FILTRO PLA

O filtro PLA é modelado em geral por um filtro AR e tem como objetivo estabelecer a necessária periodicidade na excitação do filtro PCA relacionada com o período de pitch para os segmentos vozeados. Para os segmentos não vozeados, o filtro PLA simplesmente determina a excitação do filtro PCA como combinações (no caso de mais de um coeficiente) de excitações passadas altamente correlacionadas com a excitação desejada.

Sendo $x(n)$ e $y(n)$ respectivamente os sinais de entrada e de saída, a representação do filtro PLA pode ser assim considerada:

$$y(n) = x(n) + \sum_{j=(1-T)/2}^{(T-1)/2} b_j \cdot y(n-M-j) \quad (3.8)$$

onde:

b_j são os coeficientes do filtro;

T é o número de coeficientes do filtro (ímpar);

M é o retardo do filtro.

O retardo M corresponde a um sub-múltiplo do número de amostras

correspondente ao período de pitch, ao próprio período de pitch ou a um atraso para o qual exista forte correlação entre o sinal de saída atual do filtro e o sinal de saída atrasado em M amostras.

A utilização de mais de um coeficiente permite uma interpolação caso o período de pitch não corresponda a um número inteiro de amostras [71,73-5].

Normalmente os parâmetros do filtro PLA são determinados em intervalos menores (subquadros) do que aqueles utilizados para atualização dos coeficientes do filtro PCA (quadros).

A seguir são apresentados métodos para determinação dos parâmetros do filtro PLA.

a) DETERMINAÇÃO DO RETARDO M

O retardo M do filtro PLA pode ser determinado através da correlação cruzada nomalizada entre o resíduo do filtro PCA e seus valores passados [73,76]. Em lugar do resíduo também é possível utilizar o próprio sinal de voz [17]. O ponto correspondente ao máximo da correlação cruzada normalizada para o resíduo do filtro PCA ou para o próprio sinal de voz é escolhido como o valor de M.

No caso da utilização das correlações do sinal de voz, M equivale ao valor de m que maximize a função a seguir [17]:

$$\left| \frac{\sum_{n=0}^{N-1} [s(n) \cdot s(n-m)]}{\sqrt{\sum_{n=0}^{N-1} s^2(n) \cdot \sum_{n=0}^{N-1} s^2(n-m)}} \right|, 16 \leq m \leq 160 \quad (3.9)$$

onde N é o número de amostras do subquadro.

Na procura do máximo valor da função correlação cruzada normalizada, o termo no denominador de (3.9) correspondente à energia do sinal de voz pode ser eliminado, já que é constante para qualquer valor de m .

Como a frequência correspondente ao pitch varia normalmente entre 50 e 500 Hz, a faixa de análise de M pode ser restringida ao intervalo de 16 a 160 amostras, para uma frequência de amostragem de 8 kHz.

Uma outra forma possível de determinação do retardo M consiste em calcular os coeficientes b_j do filtro PLA para todos os valores de M em uma determinada faixa. O valor ótimo de M será aquele para o qual a resposta à entrada zero do conjunto de filtros PLA e PCA conduzir à menor energia do erro ponderado entre o sinal de voz e o sinal sintetizado. Este procedimento conduz a resultados superiores em relação aqueles obtidos com o procedimento anterior, ao custo de uma maior complexidade computacional [77].

b) DETERMINAÇÃO DOS COEFICIENTES DO FILTRO PLA

Os coeficientes do filtro PLA podem ser determinados em laço aberto ou laço fechado, sendo que o último procedimento conduz a melhores resultados [77,78].

No procedimento a laço aberto, os coeficientes do filtro PLA são obtidos por técnicas LPC a partir do resíduo do filtro PCA. Neste caso, o filtro PLA inverso é excitado com o resíduo do filtro PCA, resultando:

$$t(n) = r(n) - \sum_{j=(1-T)/2}^{(T-1)/2} b_j \cdot r(n-M-j) \quad (3.10)$$

onde:

$r(n)$ é o resíduo do filtro PCA;

$t(n)$ é o resíduo do filtro PLA.

A minimização da energia E de $t(n)$ é realizada encontrando-se os valores dos b_j para os quais as derivadas parciais de E com relação aos b_j sejam nulas, o que resulta no seguinte sistema de equações lineares [79]:

$$\sum_{j=(1-T)/2}^{(T-1)/2} b_j \cdot \Phi_{rr}(M+j, M+k) = \Phi_{rr}(0, M+k) \quad , \quad \frac{1-T}{2} \leq k \leq \frac{T+1}{2} \quad (3.11)$$

onde:

$$\Phi_{rr}(M+j, M+k) = \sum_n [r(n-M-j) \cdot r(n-M-k)] \quad (3.12)$$

O intervalo de n para o cálculo de (3.12) é escolhido como sendo igual à duração de um subquadro.

No procedimento a laço fechado, os coeficientes b_j são calculados de maneira que a resposta à entrada zero dos filtros PLA e PCA conduza à menor energia do erro ponderado.

Seja $z(n)$ a resposta ponderada do filtro PCA com condições iniciais nulas quando excitado pela resposta à entrada zero do filtro PLA, $\hat{s}(n)$ o sinal de voz original menos a resposta à entrada zero do filtro PCA, ambos ponderados, e $p_o(n)$ a resposta à entrada zero do filtro ponderador de erro. O erro ponderado é dado por:

$$\hat{e}(n) = \hat{s}(n) + p_o(n) - \sum_{j=(1-T)/2}^{(T+1)/2} b_j \cdot z(n-M-j) \quad (3.13)$$

A minimização da energia E de $\hat{e}(n)$ é obtida através da determinação dos coeficientes b_j que tornam nulas as derivadas parciais de E em relação aos b_j , resultando em:

$$\sum_{j=(1-T)/2}^{(T+1)/2} b_j \cdot \Phi_{zz}(M+j, M+k) = \Phi_{zs}(M+j, 0) \quad , \frac{1-T}{2} \leq k \leq \frac{T+1}{2} \quad (3.14)$$

onde:

$$\Phi_{zz}(M+j, M+k) = \sum_n [z(n-M-j) \cdot z(n-M-k)] \quad (3.15)$$

$$\Phi_{zs}(M+j, 0) = \sum_n \{z(n-M-j) \cdot [\hat{s}(n) + p_0(n)]\} \quad (3.16)$$

O intervalo para o cálculo de (3.15) e (3.16) corresponde à duração de um subquadro (N).

Caso $M+(1-T)/2 < N$, não é válida a solução fornecida por 3.14, já que seria preciso o conhecimento da saída atual do filtro PLA para montagem do sistema. Uma aproximação que apresenta bons resultados consiste em gerar uma estimativa da saída corrente do filtro PLA simplesmente repetindo os valores passados da saída a cada M amostras, considerando $b_j = 1$ [77].

Com o objetivo de reduzir a complexidade computacional, os sistemas representados por (3.11) e (3.14) podem ser solucionados pela decomposição de Cholesky, já que implicam matrizes simétricas.

Para nenhum dos métodos anteriores é garantida a estabilidade do filtro PLA, sendo portanto necessária a verificação de estabilidade e estabilização do mesmo, caso necessário. Para o filtro PLA de 1 coeficiente, a condição suficiente para estabilidade é $|b_0| < 1$, o que é de simples verificação. Alguns autores afirmam que a utilização de $|b_0|$ pouco maior do que a unidade, em torno de 1,4 [77], não traz problemas sensíveis de estabilidade e melhora o desempenho do codificador. Filtros PLA com maior número de coeficientes necessitam de procedimentos mais complexos de verificação de estabilidade e

de estabilização. Em [76,79] são apresentados procedimentos de síntese, verificação de estabilidade e estabilização de filtros PLA de 3 coeficientes.

Foram encontradas na bibliografia consultada algumas alternativas para contornar a implementação do filtro PLA. Em [71], é proposto um único filtro que combina os efeitos dos filtros PCA e PLA, apresentando porém algumas limitações práticas. A utilização de filtros PCA de alta ordem ($OP = 50$) em estruturas do tipo "backward", onde os coeficientes do filtro não são enviados ao decodificador, pode também eliminar a necessidade de utilização do filtro PLA [46].

3.2.3 - DETERMINAÇÃO DO GANHO E DO ÍNDICE DA "CODEWORD"

O índice k da "codeword" que melhor representa a excitação é normalmente determinado de forma exaustiva, excitando-se a cascata formada pelos filtros PLA e PCA com cada candidata do "codebook" e determinando a energia do erro ponderado. Para cada "codeword" é calculado um ganho que torna mínima a energia do erro ponderado. Aquele par "codeword" e ganho que implicar menor energia do erro dentro do subquadro, é escolhido como a melhor excitação.

O ganho g aumenta virtualmente o tamanho do "codebook", sem aumentar proporcionalmente a complexidade na determinação da melhor excitação.

Seja $\hat{e}(n)$ o erro ponderado. O valor do ganho g que minimiza a energia E de $\hat{e}(n)$ é encontrado igualando-se a zero a derivada de E em relação a g . O valor de g , neste caso, é dado por:

$$g = \frac{\sum_{n=0}^{N-1} \{[(s(n) - \tilde{s}_0(n)) * p(n) + p_0(n)] \cdot [c_k(n) * l(n) * f(n) * p(n)]\}}{\sum_{n=0}^{N-1} [c_k(n) * l(n) * f(n) * p(n)]^2} \quad (3.17)$$

onde:

$\tilde{s}_0(n)$ é a resposta à entrada zero do filtro $L(z).F(z)$;

$p(n)$ é a resposta ao impulso do filtro ponderador;

$p_0(n)$ é a resposta à entrada zero do filtro ponderador;

$l(n)$ e $f(n)$ são as respostas ao impulso dos filtros PLA e PCA;

$c_k(n)$ é a k -ésima "codeword";

* representa o operador convolução linear.

A busca da melhor "codeword" é o ponto crítico na codificação CELP em termos de complexidade computacional. Algumas formas alternativas têm sido propostas para diminuir a complexidade da busca da melhor "codeword", tais como a decomposição SVD ("Singular Value Decomposition") e a busca no domínio frequência [73,80-2]. Alguns métodos chegam a resultados sub-ótimos, tais como em [83,84], onde em um primeiro passo é formado um "codebook" de tamanho reduzido com as seqüências que apresentam a menor energia de sua diferença em relação ao resíduo da cascata de filtros PLA e PCA. Outro método utilizado é a busca em várias etapas, onde o "codebook" convencional é substituído por vários "codebooks" de tamanho reduzido. A minimização da energia do erro neste caso é feita por etapas, onde cada etapa refere-se ao acréscimo da contribuição de cada "codebook" [56]. A utilização de códigos algébricos ao invés de "codebooks" estocásticos possibilita a aplicação de algoritmos mais rápidos para determinação do índice da "codeword" [85-7].

3.2.4 - PONDERAÇÃO DO ERRO

A utilização do filtro de ponderação do erro tem como objetivo a concentração do ruído do processo de codificação nas regiões dos formantes, onde ele é mascarado devido à alta energia presente nestas frequências,

melhorando subjetivamente a qualidade do sinal sintetizado [16,88].

O filtro de ponderação apresenta-se em geral na forma a seguir:

$$P(z) = \frac{F(z/\gamma)}{F(z)} \quad (3.18)$$

onde:

$F(z)$ é a função de transferência do filtro PCA;

γ é o fator de ponderação do erro.

Os valores usuais para γ situam-se no intervalo $0,8 \leq \gamma \leq 0,9$ [25].

Nota-se que para $\gamma = 1$, o filtro ponderador não tem qualquer influência, enquanto que para $\gamma = 0$, o filtro ponderador reduz-se ao inverso do filtro PCA, conformando o erro $e(n)$ de forma recíproca ao envelope espectral do sinal de voz.

3.2.5 - EXPANSÃO DA BANDA DO FILTRO PCA

A expansão da banda do filtro PCA na região dos formantes é necessária para prevenir possíveis picos não naturais no envelope espectral, decorrentes dos processos de determinação e de quantização dos coeficientes do filtro. Uma expansão de banda em torno de 15 Hz melhora a qualidade do sinal sintetizado [77].

A expansão de banda equivale a substituir o filtro $F(z)$ pelo filtro $F(z/\delta)$, onde δ é dado pela relação a seguir [77]:

$$\Delta f = \frac{-f_s}{\pi} \cdot \ln(\delta) \quad (3.19)$$

onde:

Δf é a expansão da banda em Hz;

f_s é a frequência de amostragem em Hz;

δ é o fator de expansão da banda do filtro PCA.

A Fig.3.4 apresenta a magnitude da resposta em frequência de um filtro PCA de 10 coeficientes com $\delta = 1$ e $\delta = 0,95$ (deslocado de -20 dB).

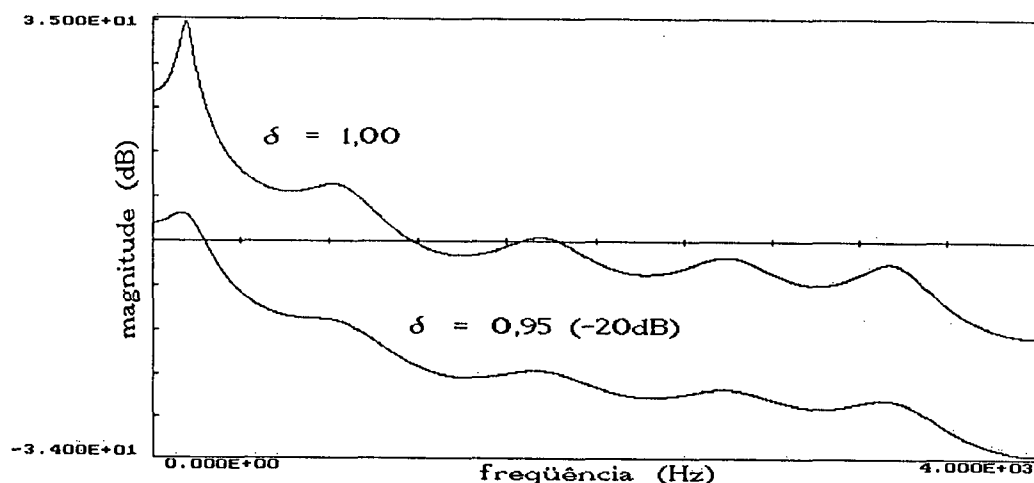


Fig.3.4 - Resposta em frequência de um filtro PCA para $\delta = 1$ e $\delta = 0,95$.

3.2.6 - PÓS-FILTRAGEM

Avaliações práticas demonstram que a qualidade do sinal sintetizado decresce rapidamente para taxas abaixo de 8 kbit/s na codificação CELP, com efeito semelhante ao da adição de um ruído ao sinal original [77]. A técnica da pós-filtragem pode ser utilizada para melhorar a qualidade subjetiva do sinal sintetizado. A pós-filtragem enfatiza os picos do espectro, aumentando a relação sinal/ruído subjetiva ao custo do aumento da distorção (relação sinal/ruído objetiva) [77].

Um pós-filtro baseado no filtro PCA e que implica a ênfase das frequências dos formantes, pode ter a forma a seguir [77,89]:

$$\frac{F(z/\beta)}{F(z/\alpha)} = \frac{1 - \sum_{i=1}^{OP} \alpha^i \cdot a_i \cdot z^{-i}}{1 - \sum_{i=1}^{OP} \beta^i \cdot a_i \cdot z^{-i}}, 0 \leq \alpha, \beta \leq 1 \quad (3.20)$$

onde:

$F(z)$ é a função de transferência do filtro PCA;

α e β são os fatores de pós-filtragem do filtro PCA;

Um pós-filtro baseado no filtro PLA de 1 coeficiente e que implica a ênfase das frequências relacionadas com o pitch é dado por [77]:

$$L(z/\epsilon^{1/M}) = \frac{1}{1 - \epsilon \cdot b_0 \cdot z^{-M}}, 0 \leq \epsilon \leq 1 \quad (3.21)$$

onde:

$L(z)$ é a função de transferência do filtro PLA;

ϵ é o fator de pós-filtragem do filtro PLA.

A pós-filtragem pode ser realizada com cada um dos filtros isoladamente ou em cascata (Fig.3.5), obtendo-se melhores resultados no último caso [77].

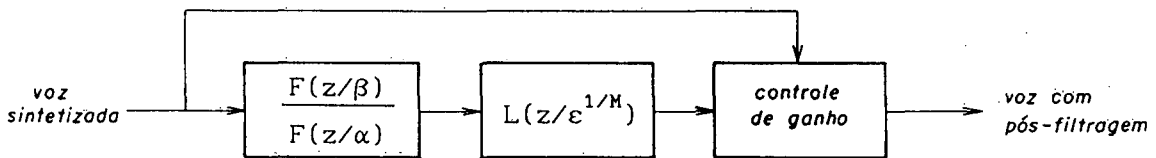


Fig.3.5 - Sistema de pós-filtragem combinado.

O procedimento de pós-filtragem apresentado introduz uma amplificação dependente do sinal de entrada, sendo necessária uma correção do ganho entre a entrada e a saída dos filtros. Esta correção pode ser realizada, por exemplo, igualando-se as energias dos sinais de entrada e de saída em um determinado

intervalo de amostras (subquadro) [77]. Para uma taxa em torno de 4800 bps, $\alpha = 0$ e $\beta = 0,3$ ou $\alpha = 0,5$ e $\beta = 0,8$ mostraram-se escolhas adequadas. O valor ótimo para ϵ situa-se em torno de 0,3 [77].

3.2.7 - CEIFAMENTO CENTRAL DO "CODEBOOK"

O ceifamento central do "codebook", consiste na eliminação das amostras de cada "codeword" que tiverem amplitudes menores em módulo do que um determinado limiar pré-estabelecido. A ocorrência de amostras nulas nas "codewords" pode implicar uma pequena redução de complexidade na filtragem das mesmas, dependendo da forma de implementação. Para um "codebook" gaussiano com variância unitária, valores em torno de $\pm 1,2$ para o ceifamento do "codebook" têm sido utilizados com sucesso [77]. Valores mais elevados podem resultar em "codewords" nulas, enquanto que valores menores podem não ter efeito significativo.

3.3 - QUANTIZAÇÃO DOS PARÂMETROS DO CODIFICADOR CELP

A quantização adequada dos parâmetros do codificador CELP é um fator de suma importância para o estabelecimento de um compromisso ótimo entre taxa e qualidade. Desta forma, deseja-se quantizar cada parâmetro com a menor quantidade possível de bits, sem que ocorra degradação sensível na qualidade do sinal sintetizado.

3.3.1 - QUANTIZAÇÃO DOS COEFICIENTES DO FILTRO PCA

A quantização dos coeficientes do filtro PCA pode ser realizada

isoladamente (quantização escalar) ou em conjunto (quantização vetorial) [90], sendo que esta última conduz a uma redução significativa na taxa ao custo do aumento de complexidade, o que compromete sua utilização em tempo real [91,92].

Um aspecto importante na quantização dos coeficientes do filtro PCA é a manutenção da estabilidade deste filtro. Tal objetivo pode ser alcançado através da quantização dos coeficientes PARCOR ou dos coeficientes de reflexão, obtidos como sub-produto no método de Durbin. Assim, no processo de quantização dos coeficientes PARCOR (k_i), é suficiente garantir a condição $-1 < k_i < 1$ para garantir a estabilidade do filtro [60].

Em [60] é mencionado que a sensibilidade espectral do filtro PCA em relação ao valor de cada PARCOR aumenta muito quando os módulos destes se aproximam da unidade. Desta forma, é necessário encontrar uma maneira de equalizar a sensibilidade espectral em relação ao valor de cada PARCOR. Tal objetivo é alcançado utilizando-se a quantização uniforme do seno inverso de cada coeficiente PARCOR, ou das razões log-área dadas a seguir [1,60,71]:

$$K_i = \ln \frac{1+k_i}{1-k_i} \quad (3.22)$$

Em [93] concluiu-se que ambas as transformações conduzem a bons resultados, com ligeira superioridade a favor da transformação do seno inverso.

Outro aspecto importante diz respeito à sensibilidade espectral em relação ao índice do PARCOR. Em [94] foi verificado que a sensibilidade espectral é decrescente com o aumento do índice do PARCOR. Desta forma, os PARCOR iniciais necessitam de mais bits para sua quantização do que os finais. Em [71] é apresentada uma proposta para a alocação de bits em função do

seno inverso dos PARCOR, bem como os limites para quantização, valores estes reproduzidos na Tabela 3.1.

Tabela 3.1 - Número de bits e limites para quantização do seno inverso dos PARCOR.

Coeficiente	Mínimo	Máximo	Bits	Coeficiente	Mínimo	Máximo	Bits
1	-1,380	0,540	5	11	-0,390	0,390	2
2	-0,300	1,230	5	12	-0,330	0,330	2
3	-1,110	0,390	4	13	-0,240	0,270	2
4	-0,300	1,110	4	14	-0,270	0,300	2
5	-0,900	0,450	3	15	-0,210	0,300	1
6	-0,300	0,900	3	16	-0,240	0,330	1
7	-0,750	0,570	3	17	-0,180	0,300	1
8	-0,450	0,600	3	18	-0,210	0,300	1
9	-0,540	0,450	3	19	-0,180	0,270	1
10	-0,450	0,360	3	20	-0,180	0,270	1

A aplicação das propriedades estatísticas dos PARCOR na quantização dos mesmos foi estudada em [94], onde os níveis de quantização foram distribuídos segundo as probabilidades de ocorrência de cada PARCOR, de forma a minimizar o erro de quantização.

A quantização dos LSP ("Line Spectrum Pairs") também tem apresentado bons resultados na quantização dos coeficientes do filtro PCA [95-7].

A quantização vetorial dos coeficientes do filtro PCA com 10 bits para 10 coeficientes tem sido utilizada com bons resultados [98], enquanto que a quantização escalar destes coeficientes pode conduzir à necessidade de até 40 bits [99].

Um método de alocação dinâmica dos bits para quantização dos coeficientes dos filtros PCA, PLA e da excitação é proposto em [100].

A interpolação das razões log-área, seno inverso dos PARCOR, LSP e coeficientes de autocorrelação conduz sempre a filtros estáveis [93].

3.3.2 - QUANTIZAÇÃO DOS PARÂMETROS DO FILTRO PLA

O parâmetro M (atraso do filtro PLA) é em geral quantizado isoladamente. Sendo este parâmetro de natureza discreta, sua quantização é facilitada. A consideração da faixa de M variando entre 16 e 160 amostras (frequência de pitch variando entre 50 Hz e 500 Hz para $f_s = 8$ kHz) implica necessidade de 8 bits para quantização deste parâmetro. Como não é necessária a utilização plena dos 8 bits, na prática tem-se reduzido a faixa de variação de M para o intervalo de 20 a 147 amostras, implicando a necessidade de 7 bits para quantização deste parâmetro, sem perda sensível de qualidade.

Com relação aos coeficientes do filtro PLA, no caso da utilização de apenas um coeficiente, tem sido adotado o procedimento de quantizar uniformemente o logaritmo do módulo deste parâmetro. A utilização de 4 ou 5 bits, incluindo o bit referente ao sinal, tem-se mostrado satisfatória [71]. Em [71] é apresentada ainda uma proposta para a quantização dos coeficientes no caso da utilização de 3 coeficientes no filtro PLA.

Na quantização dos coeficientes do filtro PLA, podem também ser utilizados procedimentos de quantização vetorial [77].

3.3.3 - QUANTIZAÇÃO DO GANHO

Durante o processo de análise, o ganho é o último parâmetro a ser quantizado, e sua determinação pode ser feita com base nos demais parâmetros já quantizados.

Em [1] é sugerido o procedimento de quantização uniforme do logaritmo do módulo do ganho. A alocação dos níveis de quantização pode ser realizada através de outros procedimentos, como o algoritmo LBG [37,101] e o quantizador

não uniforme de Lloyd-Max [77]. De maneira geral, reporta-se que 5 bits são suficientes para quantizar adequadamente o parâmetro ganho no codificador CELP [101].

3.3.4 - NÚMERO DE "CODEWORDS"

O procedimento de codificação da excitação no codificador CELP convencional é basicamente um processo de quantização vetorial, onde a melhor aproximação é determinada pela menor energia do erro ponderado. O número de bits necessários para quantizar vetorialmente a excitação está diretamente relacionado com o número de "codewords" no "codebook". Via de regra são utilizados "codebooks" de 10 bits (1024 "codewords") [24,34,102]. A utilização de "codebooks" com mais de 1024 "codewords" não traz ganhos significativos, além do que aumenta significativamente a complexidade computacional.

3.4 - ANÁLISE DO DESEMPENHO DO CODIFICADOR CELP

Com o objetivo de estabelecer padrões de referência para comparação de novas estruturas de codificação paramétrica, foi estudado o comportamento do codificador CELP frente a variações em seus parâmetros e quanto às formas de determinação e de quantização dos mesmos.

Como referência inicial, foi implementado o codificador CELP com as características propostas em [102] e apresentadas na Tabela 3.2.

No artigo da referência [102] não está explicitada a forma de determinação dos parâmetros e estabilização do filtro PLA. Foram adotadas alternativas que a priori deveriam conduzir à melhor qualidade do sinal sintetizado.

Tabela 3.2 - Parâmetros adotados como referência para o codificador CELP.

Segmento para análise LPC.....	25ms (200 amostras)
Janela de ponderação.....	Hamming
Método de determinação do filtro PCA.....	Autocorrelação
Atualização do filtro PCA (N_q).....	20ms (160 amostras)
Ordem do filtro PCA (OP).....	10
Fator de ponderação do erro (γ).....	0,900
Fator de expansão da banda do filtro PCA (δ)...	0,994 (15Hz)
Atualização da excitação (N).....	5ms (40 amostras)
Número de coeficientes do filtro PLA (T).....	$1(b_i \leq 1)$
Forma de determinação do filtro PLA.....	Laço fechado/M ótimo
Faixa do atraso M.....	20...147
Número de "codewords" (K).....	1024
Ceifamento central do "codebook".....	$\pm 1,3$

Nesta fase de análise, optou-se pela utilização do modelo não quantizado, de forma que toda degradação nos resultados obtidos fosse devida unicamente às restrições impostas pelos parâmetros adotados no modelo.

Todas as simulações foram executadas com os arquivos denominados por AM e AF, correspondentes respectivamente a um segmento de voz masculina e um segmento de voz feminina, contendo uma elocução com aproximadamente 15 segundos, conforme especificado no Apêndice C. O processamento foi realizado em uma estação de trabalho IBM da linha RISC 6000, utilizando-se linguagem C.

A Tabela 3.3 apresenta os resultados objetivos obtidos com a utilização dos parâmetros da Tabela 3.2, em termos da Relação Sinal/Ruído Segmentada (RSRSEG), conforme definida em [11] e calculada com blocos de 40 amostras, além do tempo de processamento aproximado (T_p) em minutos.

Tabela 3.3 - Resultados objetivos obtidos como referência inicial para o codificador CELP.

Arquivo	Duração(s)	RSRSEG(dB)	T_p (min.)
AM/AF	13,96/14,73	12,55/11,82	40/43

• Análise subjetiva informal: Os resultados obtidos como referência inicial revelaram excelente qualidade, com degradação mínima em relação aos arquivos originais. Foi observada uma pequena redução de agudos, limitação esta característica do codificador CELP convencional e mencionada em [102].

3.4.1 - VARIAÇÃO DA ORDEM DO FILTRO PCA

A Tabela 3.4 apresenta os resultados objetivos obtidos em função da variação do número de coeficientes do filtro preditor de curto atraso. A Fig.3.6 apresenta os mesmos resultados em forma de gráfico.

O comportamento da RSRSEG com o aumento do número de coeficientes do filtro PCA revela uma tendência à saturação para ordens próximas à 14. Deve ser ressaltado que a utilização da expansão de banda contribui para o decréscimo da RSRSEG com ordens maiores do que 14, já que pondera de forma exponencialmente decrescente os coeficientes do filtro PCA.

Tabela 3.4 - Resultados objetivos em função da variação da ordem do filtro PCA no codificador CELP.

Arquivo	Ordem do PCA	RSRSEG(dB)	Tp(min.)
AM/AF	0	6,94/ 7,43	26/26
AM/AF	1	10,69/10,12	26/26
AM/AF	2	11,75/10,97	27/28
AM/AF	4	12,11/11,36	31/32
AM/AF	6	12,09/11,45	34/38
AM/AF	8	12,10/11,70	37/40
AM/AF	10	12,55/11,82	40/42
AM/AF	12	12,64/11,93	43/46
AM/AF	14	12,69/11,96	46/48
AM/AF	16	12,62/11,88	48/51
AM/AF	20	12,65/11,84	53/56
AM/AF	30	12,55/11,81	60/63
AM/AF	40	12,60/11,61	63/66

• Análise subjetiva informal: Para ordens superiores a 14, os resultados obtidos são praticamente idênticos quanto à qualidade. A utilização de 8 coeficientes revelou um resultado ruim com o arquivo AM e bom com arquivo AF, indicando que a generalização de tal condição não é adequada. A utilização de 12 coeficientes apresentou resultados de muito boa qualidade para ambos arquivos, embora a diferença em relação ao uso de 10 coeficientes não tenha sido muito significativa.

Todas as simulações seguintes foram realizadas com ordem do filtro preditor de curto atraso igual a 10.

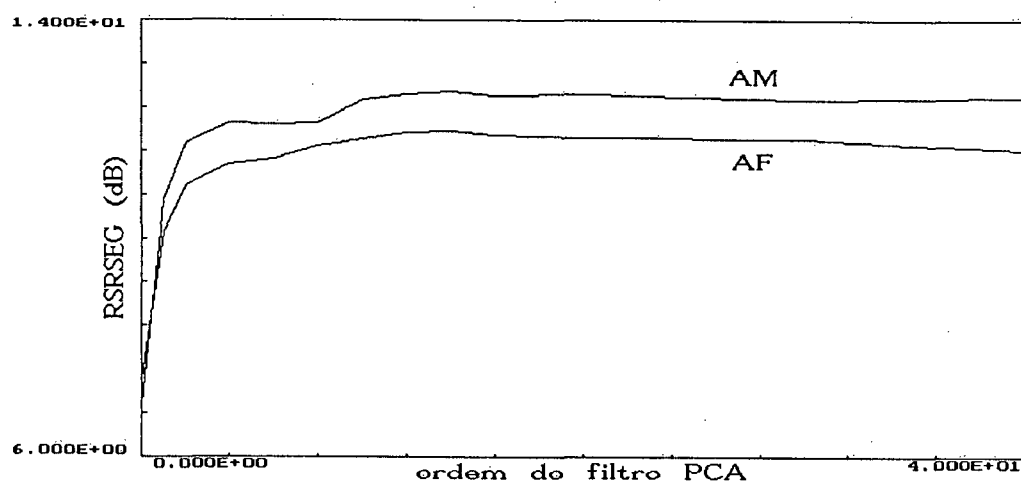


Fig.3.6 - Relação sinal/ruído segmentada em função da ordem do filtro PCA.

3.4.2 - VARIAÇÕES NO JANELAMENTO DO SINAL DE VOZ

Neste item foi estudado o comportamento do codificador CELP em função da variação do tipo de janela e do número de amostras utilizadas para análise LPC, com manutenção do intervalo de atualização dos coeficientes do filtro PCA em 160 amostras. A Tabela 3.5 apresenta os resultados objetivos obtidos com a

utilização das janelas de Hamming e retangular e variação do número de amostras para análise LPC.

Tabela 3.5 - Resultados objetivos em função do janelamento do sinal de voz no codificador CELP.

Arquivo	Janela	Número de Amostras	RSRSEG(dB)	Tp(min.)
AM/AF	Hamming	160	12,56/11,78	40/42
AM/AF	Hamming	200	12,55/11,82	40/42
AM/AF	Hamming	240	12,53/11,83	40/42
AM/AF	Hamming	280	12,47/11,80	40/42
AM/AF	Hamming	320	12,48/11,82	40/43
AM/AF	Retangular	160	12,67/11,92	40/42
AM/AF	Retangular	200	12,52/11,86	40/42
AM/AF	Retangular	240	12,57/11,88	40/42
AM/AF	Retangular	280	12,43/11,84	40/42
AM/AF	Retangular	320	12,48/11,77	40/43

● Análise subjetiva informal: A utilização da janela de Hamming resultou em uma voz mais clara do que a obtida com a janela retangular. A diferença é facilmente percebida, não sendo contudo muito acentuada. Com o aumento do número de amostras na análise LPC, não foi percebida qualquer melhora subjetiva, no caso da janela retangular. No caso da janela de Hamming, foi percebida uma ligeira melhora na qualidade do sinal sintetizado com o aumento do intervalo de análise até 240 amostras. Para intervalos maiores foram percebidas pequenas degradações, relacionadas com o problema da variação das estatísticas do sinal para longos intervalos de tempo (sinal não estacionário).

Os melhores resultados subjetivos para os dois segmentos de voz analisados foram obtidos com a utilização da janela de Hamming e de um intervalo de análise LPC de 240 amostras, para um intervalo de 160 amostras na atualização dos coeficientes do filtro PCA. Todas as simulações posteriores foram realizadas considerando tais condições.

3.4.3 - UTILIZAÇÃO DA MEMÓRIA INICIAL DOS FILTROS PCA E PONDERADOR DE ERRO

A memória inicial dos filtros PCA, decorrente do sinal sintetizado nos quadros anteriores (resposta à entrada zero) é a única componente do sinal sintetizado que não é controlada por algum parâmetro dentro do subquadro de análise. Neste item, foi estudado o efeito da eliminação da memória inicial dos filtros PCA e ponderador de erro entre quadros e subquadros, o que resultaria em uma diminuição da complexidade, ainda que pequena, nos modelos de análise e síntese.

A Tabela 3.6 apresenta os resultados objetivos obtidos com a eliminação da memória inicial do filtro PCA entre quadros e subquadros e da memória inicial do filtro ponderador de erro entre subquadros, nas etapas de análise e síntese. Pode ser observado que a redução da complexidade computacional não é significativa em qualquer uma das opções.

Tabela 3.6 - Resultados objetivos com utilização ou não de memória inicial nos filtros PCA e ponderador de erro.

Arquivo	Condição	RSRSEG(dB)	Tp(min.)
AM/AF	C/memória	12,53/11,83	40/42
AM/AF	S/memória PCA entre quadros	11,47/10,91	40/42
AM/AF	S/memória PCA entre subquadros	9,46/ 9,24	40/42
AM/AF	S/mem. ponderador entre subquadros	12,74/12,10	40/42

- **Análise subjetiva informal:** A não utilização da memória inicial do filtro PCA entre quadros resultou em um sinal sintetizado sensivelmente ruidoso, porém bem inteligível. A degradação foi muito mais acentuada no caso da não utilização da memória entre subquadros, resultando em uma qualidade inaceitável. A não utilização da memória do filtro ponderador de erro, embora

aumentando a RSRSEG, resultou em uma pequena redução na qualidade subjetiva em relação ao resultado obtido com utilização da memória.

Como conclusão, tem-se que a utilização da memória inicial do filtro PCA, tanto nas transições de quadros, como nas transições de subquadros, é indispensável para a obtenção de um sinal sintetizado de boa qualidade. Já a utilização ou não da memória do filtro ponderador de erro, embora também influenciando na qualidade do sinal sintetizado, não é tão crítica.

Todas as simulações a seguir foram realizadas incluindo-se os efeitos das memórias iniciais dos filtros PCA e ponderador de erro nas transições de quadros e subquadros.

3.4.4 - MÉTODO DE DETERMINAÇÃO DOS COEFICIENTES DO FILTRO PCA

A Tabela 3.7 apresenta os resultados objetivos obtidos com a utilização dos métodos da autocorrelação, covariância e covariância estabilizado pela limitação dos PARCOR entre -1 e 1, na obtenção dos coeficientes do filtro PCA.

Tabela 3.7 - Resultados objetivos com diversos métodos para determinação dos coeficientes do filtro PCA.

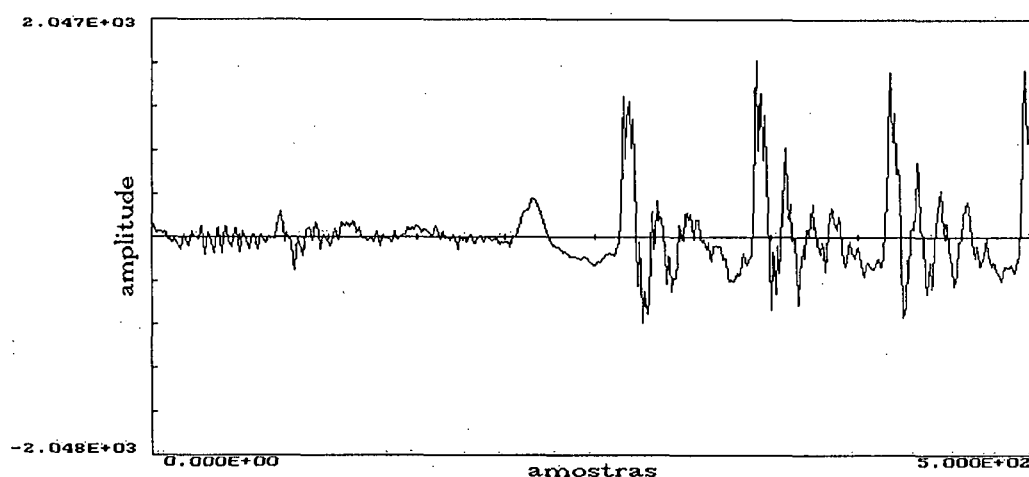
Arquivo	Método	RSRSEG(dB)	Tp(min.)
AM/AF	Autocorrelação	12,53/11,83	40/42
AM/AF	Covariância	12,12/11,72	40/42
AM/AF	Covariância estabilizado	12,30/11,71	40/42

● **Análise subjetiva informal:** De uma forma geral, a única diferença percebida na qualidade do sinal sintetizado, utilizando-se o método da autocorrelação ou o método da covariância, foi a ocorrência de fortes ruídos localizados, devidos à síntese com filtros instáveis. O processo de

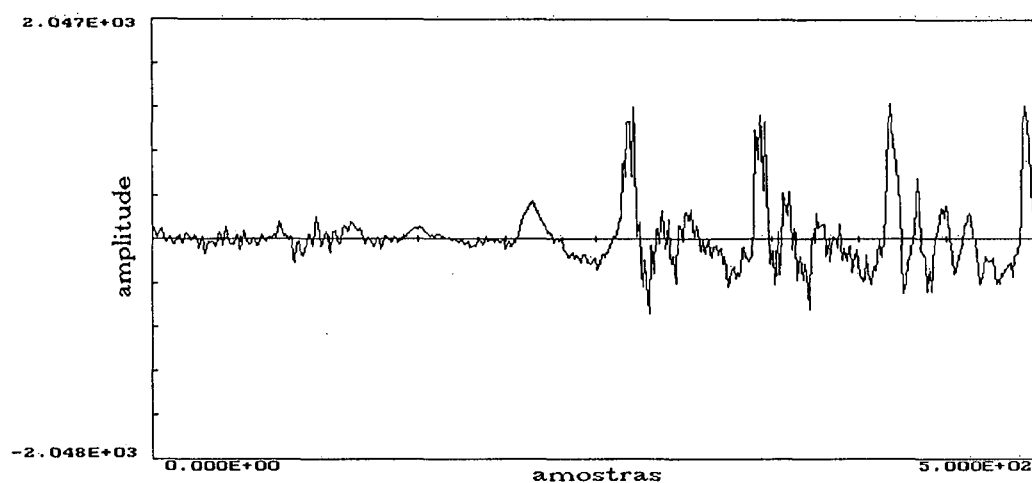
estabilização utilizado não eliminou integralmente o problema observado com o método da covariância, provavelmente porque o filtro estável gerado tem características espectrais não muito próximas das desejadas.

A Fig. 3.7 apresenta a comparação de um segmento de voz com os segmentos sintetizados pelos métodos da autocorrelação e covariância.

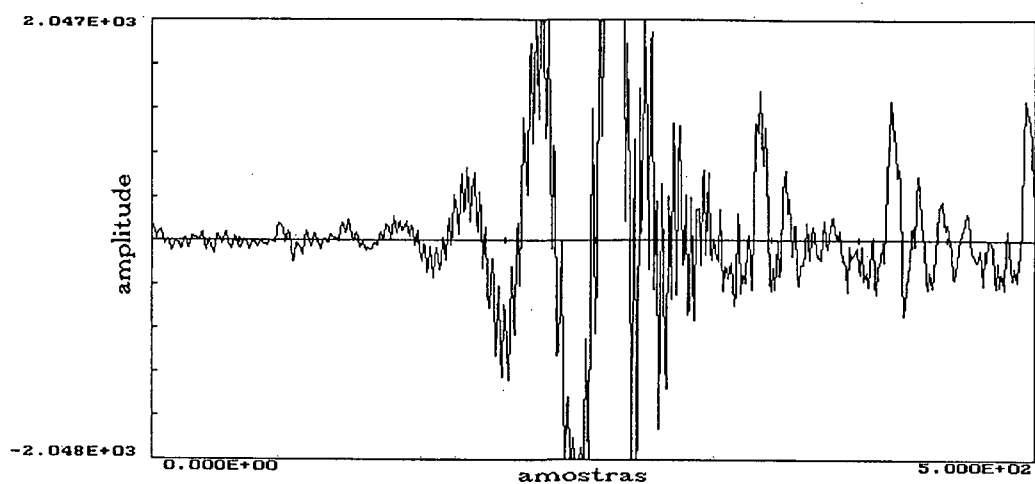
Como conclusão, tem-se que o método da autocorrelação constitui uma escolha adequada pela menor complexidade, qualidade ao mesmo nível do método da covariância e ausência de problemas decorrentes da síntese com filtro PCA



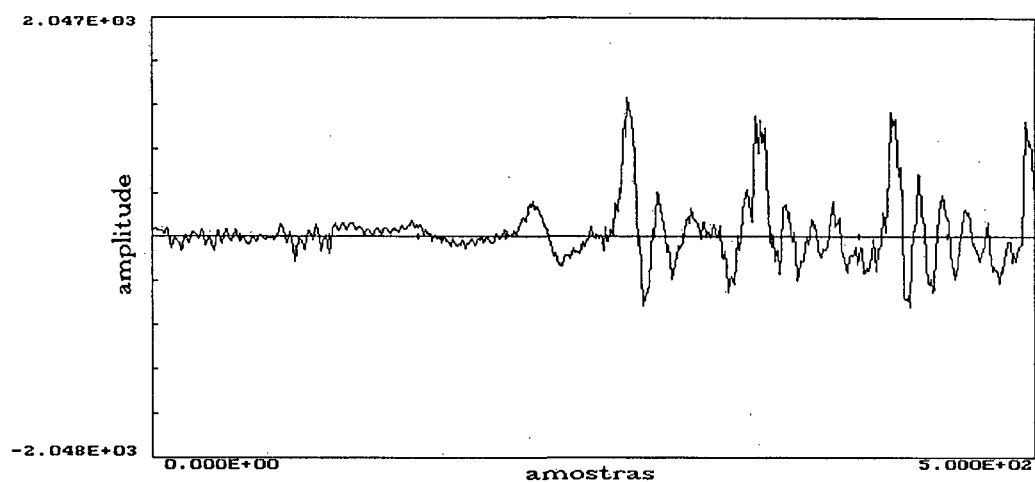
- segmento de voz original



- segmento de voz obtido pelo método da autocorrelação



- segmento de voz obtido pelo método da covariância com filtro instável



- segmento de voz obtido com o método da covariância estabilizado

Fig.3.7 - Segmentos de voz obtidos com os métodos da autocorrelação e covariância.

instável. Para as simulações posteriores foi adotado o método da autocorrelação.

3.4.5 - VARIAÇÃO DO INTERVALO DE ATUALIZAÇÃO DOS COEFICIENTES DO FILTRO PCA

Neste item foi avaliada a variação da qualidade do sinal de voz

sintetizado em função da variação do número de amostras para atualização dos coeficientes do filtro PCA (duração do quadro), mantendo-se fixa a duração dos subquadros. Os resultados objetivos são apresentados na Tabela 3.8.

Tabela 3.8 - Resultados objetivos em função da variação do quadro de análise.

Arquivo	Quadro(amostras)	RSRSEG(dB)	Tp(min.)
AM/AF	40	13,06/12,47	40/42
AM/AF	80	12,89/12,17	40/42
AM/AF	160	12,53/11,83	40/42
AM/AF	240	12,30/11,65	40/42
AM/AF	320	12,22/11,57	40/42

● Análise subjetiva informal: A variação do intervalo de atualização dos coeficientes LPC entre 40 e 320 amostras (5 e 40 ms) apresentou resultados subjetivamente muito semelhantes, todos de boa qualidade.

Para as simulações a seguir, foi adotado o intervalo de 160 amostras (20 ms) para atualização dos coeficientes LPC, reservando os intervalos maiores para aplicações onde seja necessária uma redução mais significativa da taxa, sem grande comprometimento da qualidade.

3.4.6 - MÉTODO DE DETERMINAÇÃO DO RETARDO DO FILTRO PLA

A Tabela 3.9 apresenta os resultados objetivos obtidos com os procedimentos de determinação ótima de M (busca exaustiva) e determinação não ótima (máxima correlação do resíduo do filtro PCA), para $20 \leq M \leq 147$.

Apesar da maior complexidade computacional, o procedimento de determinação do M ótimo foi mantido para as simulações posteriores, em vista da melhor qualidade apresentada nos resultados.

Tabela 3.9 - Resultados objetivos em função do método de determinação do retardo do filtro PLA.

Arquivo	Método	RSRSEG(dB)	Tp(min.)
AM/AF	M ótimo	12,53/11,83	40/42
AM/AF	M não ótimo	11,11/10,41	35/37

• Análise subjetiva informal: A determinação do M não ótimo resultou em um ligeiro decréscimo na qualidade do sinal de voz sintetizado. Se a redução de complexidade for um objetivo prioritário, este procedimento pode ser uma alternativa interessante.

3.4.7 - VARIAÇÃO DA FAIXA DO RETARDO DO FILTRO PLA

O desempenho do codificador CELP foi analisado em função da faixa do retardo M do filtro PLA dentro do intervalo de 16 a 160 amostras, o que equivale a uma faixa de pitch entre 50 Hz e 500 Hz. Os resultados objetivos são apresentados na Tabela 3.10.

Tabela 3.10 - Resultados objetivos em função da faixa do retardo do filtro PLA.

Arquivo	Faixa para M	RSRSEG(dB)	Tp(min.)
AM/AF	16... 79	11,58/11,44	37/39
AM/AF	16...143	12,58/11,81	40/42
AM/AF	20...147	12,53/11,83	40/42
AM/AF	33...160	12,66/11,88	40/42
AM/AF	16...160	12,66/11,89	41/43

• Análise subjetiva informal: A alternativa com $16 \leq M \leq 79$ (6 bits) apresentou uma leve degradação em relação à alternativa com $16 \leq M \leq 160$,

sendo entretanto o resultado de boa qualidade. Nos demais casos, a qualidade subjetiva foi praticamente idêntica à alternativa $16 \leq M \leq 160$.

Para as simulações posteriores, foi adotada a faixa $33 \leq M \leq 160$ que, apesar de cobrir uma faixa de pitch menor em frequência do que a alternativa convencional $20 \leq M \leq 147$, apresentou uma melhor RSRSEG do que esta última para os dois segmentos de voz analisados.

3.4.8 - MÉTODO DE DETERMINAÇÃO DOS COEFICIENTES DO FILTRO PLA

Foram comparados neste item os procedimentos em malha aberta e malha fechada para determinação dos coeficientes do filtro PLA. Os resultados objetivos são apresentados na Tabela 3.11.

Tabela 3.11 - Resultados objetivos em função do método de determinação dos coeficientes do filtro PLA.

Arquivo	Método de Cálculo	RSRSEG(dB)	Tp(min.)
AM/AF	Malha aberta	11,48/10,66	38/40
AM/AF	Malha fechada	12,66/11,88	40/42

- **Análise subjetiva informal:** Ambos os procedimentos conduziram a resultados muito bons, com uma ligeira superioridade do procedimento em malha fechada.

Como alternativa na redução da complexidade, o procedimento em malha aberta, apesar da qualidade ligeiramente inferior, pode ser uma alternativa interessante.

Para as simulações posteriores foi mantido o procedimento em malha fechada para determinação do PLA, por ter apresentado melhores resultados.

3.4.9 - VARIAÇÃO DO NÚMERO DE COEFICIENTES DO FILTRO PLA

A influência do número de coeficientes do filtro PLA no desempenho do codificador CELP foi avaliada e os resultados são apresentados na Tabela 3.12.

Com relação à eliminação do filtro PLA, cabe ressaltar que o tempo de processamento apresentado refere-se à determinação de um filtro PLA de 1 coeficiente e posterior cancelamento deste coeficiente ($b_j = 0$).

Tabela 3.12 - Resultados objetivos em função do número de coeficientes do filtro PLA.

Arquivo	Coeficientes do PLA	RSRSEG(dB)	Tp(min.)
AM/AF	0	8,49/ 7,75	40/42
AM/AF	1	12,66/11,88	40/42
AM/AF	3	13,08/12,49	46/49

- Análise subjetiva informal: A eliminação do filtro PLA implicou em degradação acentuada no sinal sintetizado. Já a utilização de 3 coeficientes apresentou uma ligeira melhora em relação à utilização de 1 coeficiente.

Concluiu-se nesta análise que a utilização de um filtro PLA de 3 coeficientes não se justifica em vista dos aumentos de complexidade e taxas decorrentes. Devido a este fato e considerando que a eliminação do filtro PLA conduz à degradações acentuadas, a utilização do filtro PLA de um coeficiente é a escolha mais adequada e foi adotada nas demais simulações.

3.4.10 - ESTABILIZAÇÃO DO FILTRO PLA

Neste item foi estudada a limitação do coeficiente do filtro PLA de

maneira que os problemas decorrentes da síntese com filtros instáveis não degradem sensivelmente o sinal sintetizado. A Tabela 3.13 apresenta os resultados objetivos obtidos.

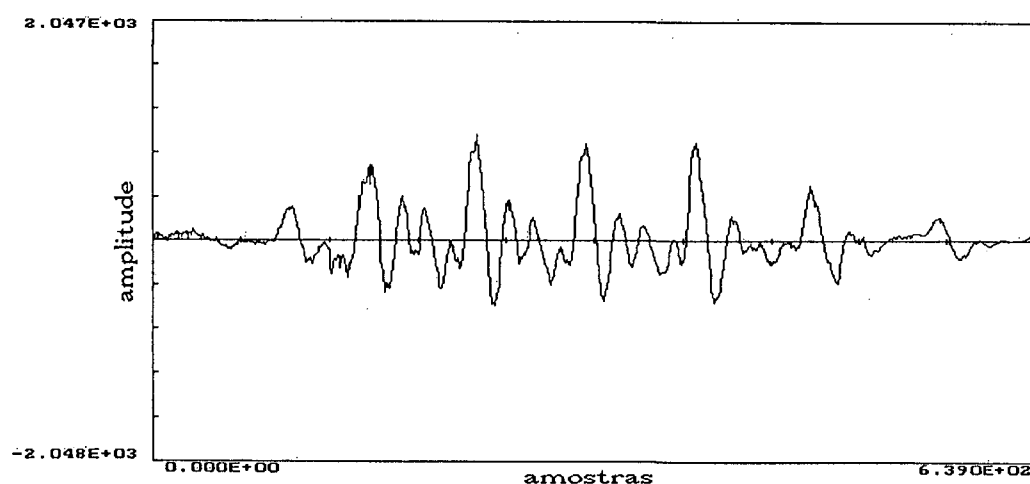
Tabela 3.13 - Resultados objetivos obtidos com a limitação de $|b_0|$.

Arquivo	Limite para $ b_0 $	RSRSEG(dB)	Tp(min.)
AM/AF	∞	12,72/11,82	40/42
AM/AF	2,0	12,73/11,90	40/42
AM/AF	1,5	12,67/11,94	40/42
AM/AF	1,0	12,66/11,88	40/42

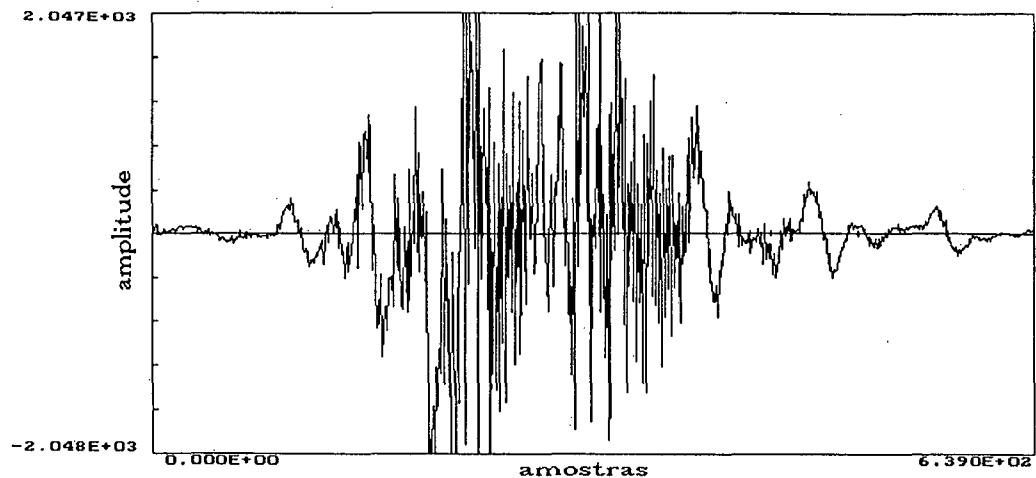
• Análise subjetiva informal: A utilização de filtros PLA instáveis ($|b_0| > 1$) resultou em um ligeiro aumento do ruído de fundo nos segmentos de voz sintetizados.

A não imposição de limites adequados para $|b_0|$ pode resultar em severas degradações no sinal sintetizado. A Fig.3.8 apresenta segmentos de voz sintetizados obtidos com filtros PLA estáveis e instáveis.

Como conclusão, tem-se que a utilização de filtros PLA estáveis e



- segmento de voz obtido com filtro PLA não instável ($|b_0| \leq 1$)



- segmento de voz obtido com filtro PLA instável ($|b_0| < \infty$)

Fig.3.8 - Segmentos de voz obtidos com filtros PLA não instáveis e instáveis.

marginalmente estáveis ($|b_0| \leq 1$) para todos os subquadros, conduz a uma melhor qualidade do sinal sintetizado, sendo portanto o procedimento mais adequado. Tal procedimento foi mantido para as simulações posteriores.

3.4.11 - CEIFAMENTO CENTRAL DO "CODEBOOK"

A idéia do ceifamento central do "codebook" proposta em [77] foi avaliada neste item. A Tabela 3.14 apresenta os resultados objetivos obtidos para vários níveis de ceifamento central de cada "codeword".

Tabela 3.14 - Resultados objetivos para vários níveis de ceifamento central do "codebook".

Arquivo	Nível de Ceifamento	RSRSEG(dB)	Tp(min.)
AM/AF	0,0	12,52/11,80	40/42
AM/AF	±0,5	12,54/11,82	40/42
AM/AF	±1,0	12,55/11,87	40/42
AM/AF	±1,3	12,66/11,88	40/42

• Análise subjetiva informal: Os resultados não apresentaram qualquer diferença de qualidade que pudesse ser percebida.

Em vista do melhor desempenho em termos objetivos e de uma possível redução na complexidade, foi adotado o nível de $\pm 1,3$ na ceifagem central do "codebook" para as simulações posteriores. A utilização de níveis maiores do que $\pm 1,3$ resultou em "codewords" nulas, não sendo portanto adequado ao "codebook" utilizado.

3.4.12 - VARIAÇÃO DA DURAÇÃO DOS SUBQUADROS

A Tabela 3.15 apresenta os resultados objetivos em função da variação do número de amostras dos subquadros.

Tabela 3.15 - Resultados objetivos em função do número de amostras do subquadro.

Arquivo	Subquadro(amostras)	RSRSEG(dB)	Tp(min.)
AM/AF	10	21,06/20,26	37/39
AM/AF	20	16,31/15,47	39/41
AM/AF	40	12,66/11,88	40/43
AM/AF	80	8,97/ 8,78	40/43
AM/AF	160	6,00/ 6,09	41/44

• Análise subjetiva informal: A melhora na qualidade subjetiva é sensível com a redução do número de amostras dos subquadros. Para subquadros de 10 amostras, a voz sintetizada foi praticamente idêntica à voz original. Com o aumento do número de amostras, a qualidade degradou-se sensivelmente. Subquadros de 160 amostras implicaram degradações elevadas.

Para as simulações posteriores, foram mantidos os subquadros de 40

amostras, reservando a utilização de subquadros maiores para redução da taxa sem aumento significativo da degradação.

3.4.13 - QUANTIZAÇÃO DOS PARÂMETROS

O estudo da quantização de cada parâmetro foi realizado em duas fases. Na primeira fase, foram determinados de forma estatística os limites de quantização do parâmetro estudado e avaliada sua influência na qualidade do sinal sintetizado através de uma quantização com infinitos níveis. Na segunda fase, foi variado o número de bits na quantização do parâmetro, analisando-se os resultantes obtidos.

A seguir são apresentados os resultados obtidos com a quantização dos parâmetros em cada uma das etapas do codificador CELP.

a) FILTRO PREDITOR DE CURTO ATRASO

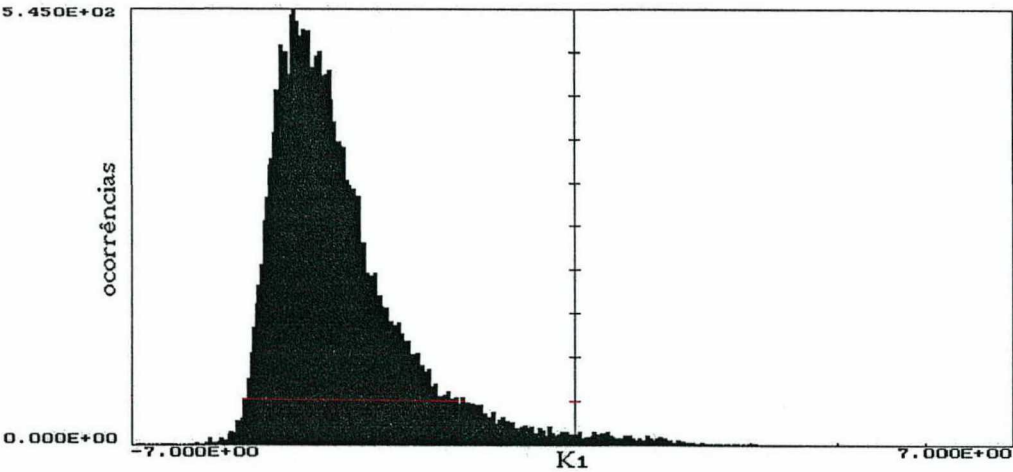
Na quantização dos PARCOR, foram avaliados os métodos de dessensibilização do seno inverso dos PARCOR e das razões log-área na quantização escalar destes parâmetros.

Os limites para quantização de cada coeficiente PARCOR foram determinados através da análise das ocorrências destes parâmetros para os arquivos de voz mencionados no Apêndice C. Os arquivos AM e AF não foram incluídos na estatística. A Tabela 3.16 apresenta os limites inferiores e superiores para o seno inverso dos PARCOR e razões log-área, obtidos estatisticamente de maneira a incluir 99% das ocorrências de cada parâmetro.

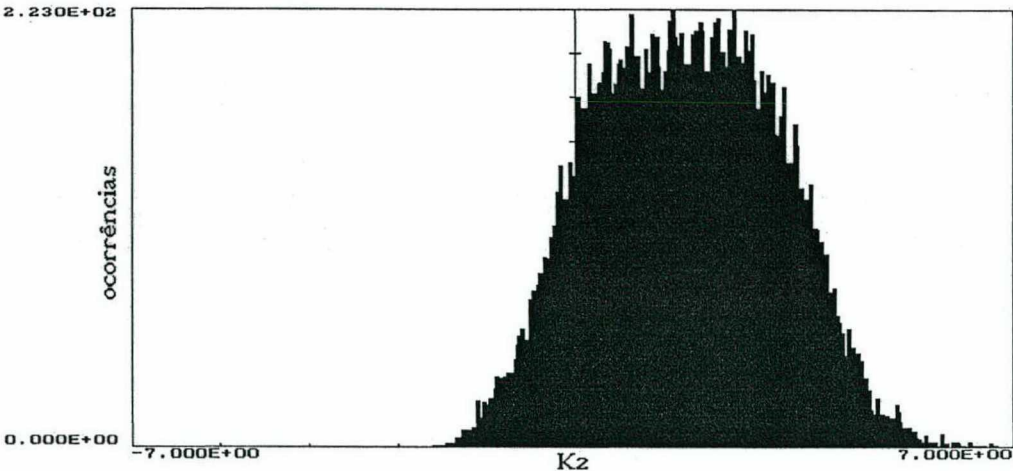
Os histogramas das razões log-área, para filtros PCA de 10 coeficientes, são apresentados na Fig.3.9.

Tabela 3.16 - Limites de quantização para o seno inverso dos PARCOR e razões log-área.

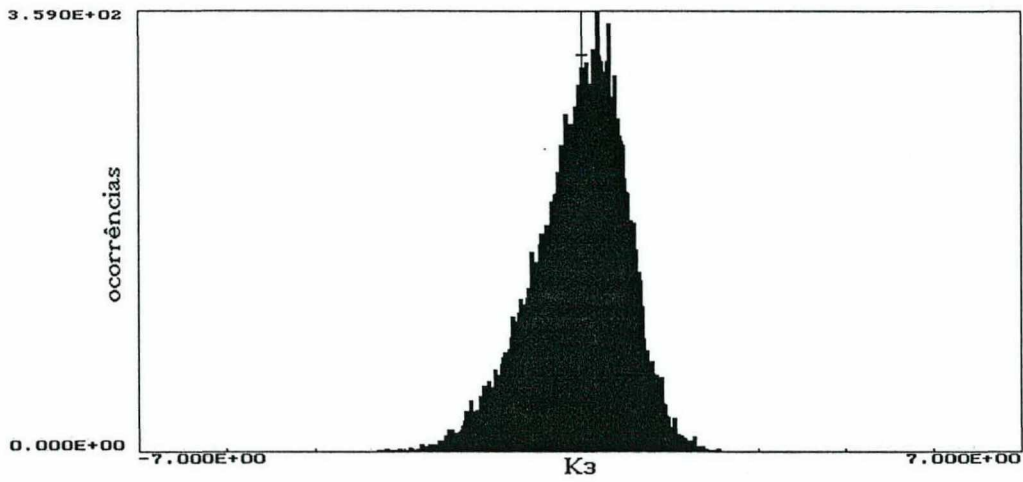
Coeficiente	Seno Inverso dos PARCOR		Razões Log-Área	
	máximo	mínimo	máximo	mínimo
1	1,434054	-0,537393	1,183514	-5,373020
2	0,673049	-1,432633	5,206196	-1,512808
3	0,905509	-0,708275	1,562927	-2,152466
4	0,422922	-1,060837	2,704809	-0,893341
5	0,777100	-0,528493	1,106387	-1,751903
6	0,381901	-0,793056	1,759764	-0,809615
7	0,749794	-0,398252	0,813543	-1,665116
8	0,425151	-0,661121	1,423698	-0,883242
9	0,690246	-0,457162	0,938898	-1,502833
10	0,337764	-0,623702	1,341559	-0,690563



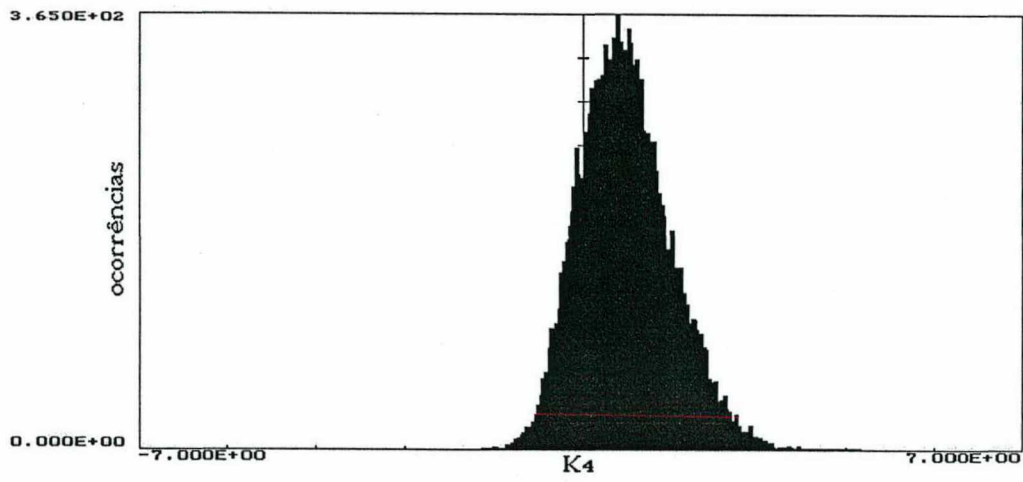
- ocorrências da razão log-área 1



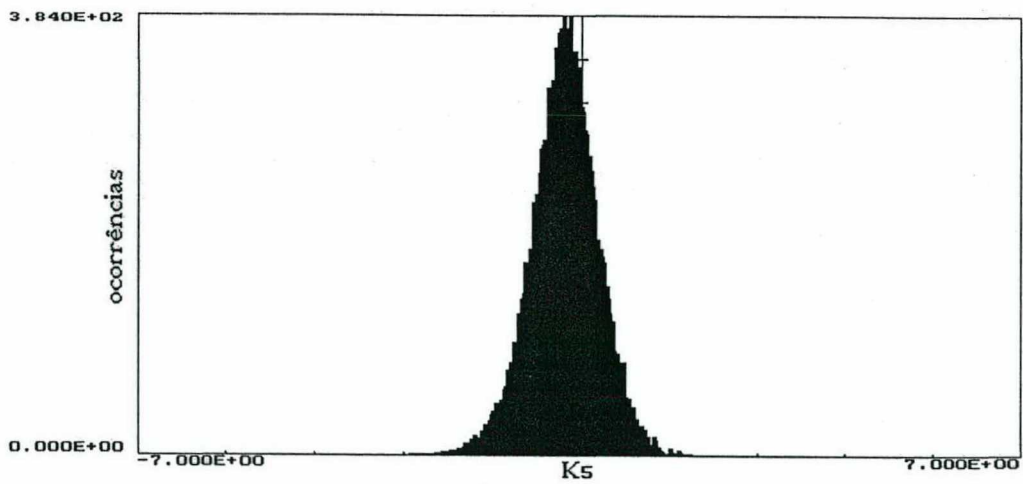
- ocorrências da razão log-área 2



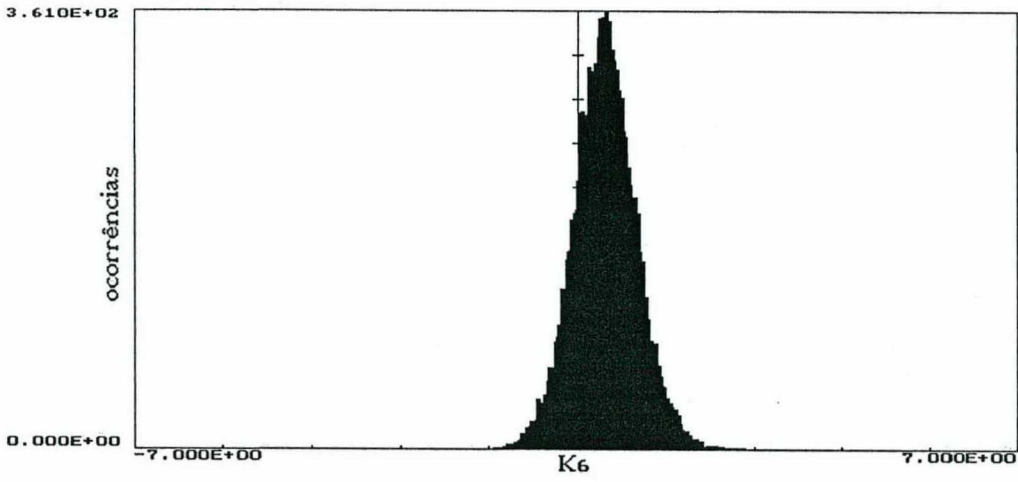
- ocorrências da razão log-área 3



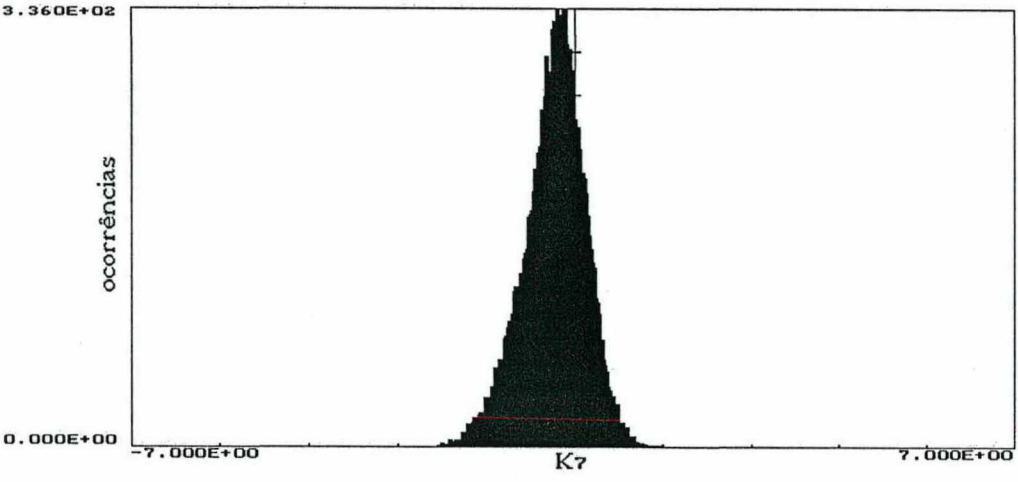
- ocorrências da razão log-área 4



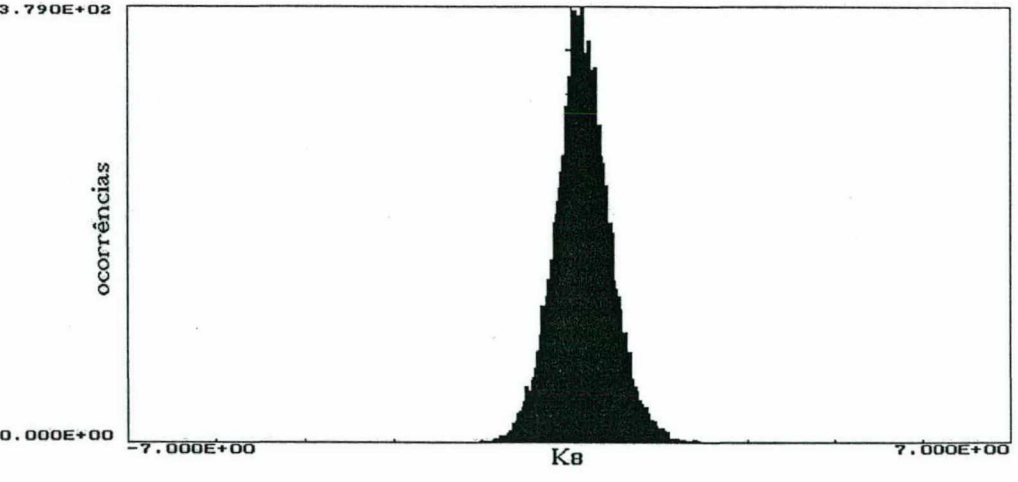
- ocorrências da razão log-área 5



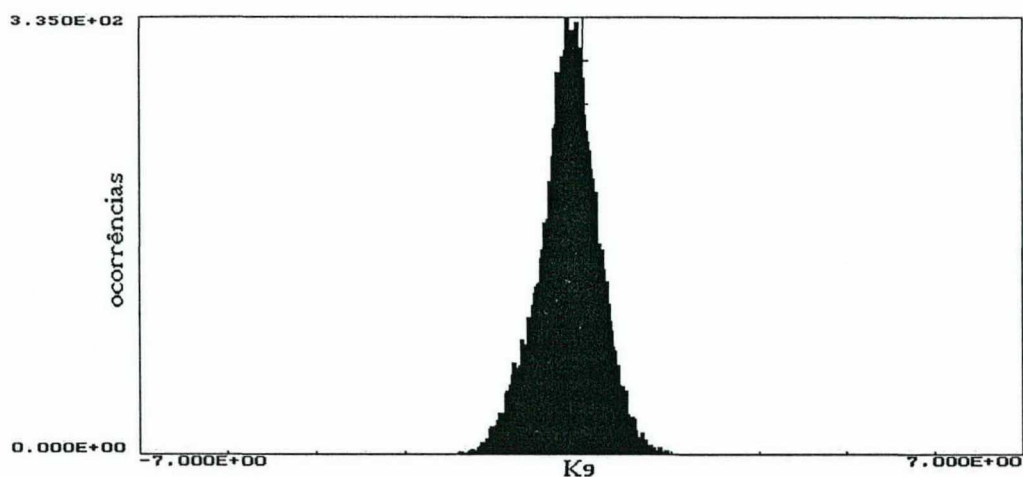
- ocorrências da razão log-área 6



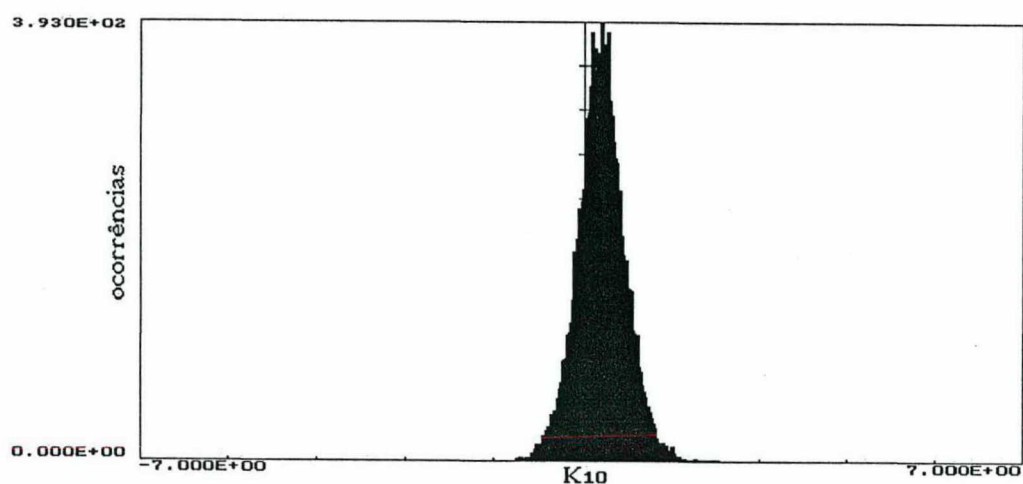
- ocorrências da razão log-área 7



- ocorrências da razão log-área 8



- ocorrências da razão log-área 9



- ocorrências da razão log-área 10

Fig.3.9 - Histogramas das razões log-área.

Para avaliar o nível de degradação introduzido pela limitação na quantização dos parâmetros, foram simuladas quantizações com infinitos níveis utilizando os limites da Tabela 3.16. Os resultados objetivos destas simulações são apresentados na Tabela 3.17.

Com relação aos resultados objetivos, as diferenças encontradas não são relevantes, indicando que os limites adotados não introduzem degradações significativas.

Tabela 3.17 - Resultados objetivos com a limitação e quantização infinita dos parâmetros do filtro PCA.

Arquivo	Procedimento de Quantização	RSRSEG(dB)	Tp(min.)
AM/AF	Sem limitação	12,66/11,88	40/42
AM/AF	Seno inverso c/limitação	12,66/11,87	40/42
AM/AF	Razões log-área c/limit.	12,61/11,88	40/42

● Análise subjetiva informal: Não foram observadas quaisquer diferenças entre os resultados com e sem limitação dos parâmetros do filtro PCA.

Os resultados objetivos utilizando-se ambos os procedimentos de quantização e a alocação de bits apresentada na Tabela 3.18 [71] estão relacionados na Tabela 3.19.

Tabela 3.18 - Bits para quantização dos parâmetros do filtro PCA.

Coeficiente	1	2	3	4	5	6	7	8	9	10
Bits	5	5	4	4	3	3	3	3	3	3

Tabela 3.19 - Resultados objetivos com a quantização do seno inverso dos PARCOR e das razões log-área.

Arquivo	Transformação	RSRSEG(dB)	Tp(min.)
AM/AF	Seno inverso	12,26/11,54	40/42
AM/AF	Logarítmica	12,29/11,66	40/42

● Análise subjetiva informal: Foram verificados resultados praticamente idênticos para ambos os procedimentos em relação àqueles obtidos com quantização infinita.

Em vista da pequena margem objetiva obtida com a quantização das razões

log-área (Tabela 3.19), foi mantido este procedimento para as demais simulações.

A Tabela 3.20 apresenta os resultados objetivos obtidos com a variação do número de bits na quantização de cada razão log-área.

Tabela 3.20 - Resultados objetivos em função do número de bits na quantização das razões log-área.

Arquivo	Alocação de Bits	RSRSEG(dB)	Tp(min.)
AM/AF	5,5,5,4,4,4,4,3,3,3	12,41/11,72	40/42
AM/AF	5,5,4,4,3,3,3,3,3,3	12,29/11,66	40/42
AM/AF	5,5,4,4,3,3,3,3,2,2	12,19/11,51	40/42
AM/AF	5,4,4,4,3,3,3,2,2,2	12,04/11,37	40/42
AM/AF	4,4,4,4,3,3,3,3,2,2	11,99/11,37	40/42

• Análise subjetiva informal: Foram verificados resultados de boa qualidade para todas as configurações de bits testadas, tendo sido percebida apenas uma pequena degradação na última em relação às demais.

Para as simulações posteriores foi adotada a configuração (5,4,4,4,3,3,3,2,2,2) por ter apresentado resultado subjetivo ao nível das demais e por conduzir a uma menor taxa (32 bits/quadro).

b) FILTRO PREDITOR DE LONGO ATRASO

O parâmetro M do filtro PLA é de natureza discreta e o número de bits necessários para sua quantização é estabelecido diretamente pela determinação dos limites de variação de M . Assim, a quantização do parâmetro M variando entre 33 e 160, é realizada com 7 bits.

Com relação ao coeficiente b_0 , no caso do filtro PLA de 1 coeficiente,

foi estudada a quantização uniforme do logaritmo natural do módulo deste parâmetro com diversos números de bits.

Os limites para a quantização de b_0 foram determinados através da análise das ocorrências para os arquivos relacionados no Apêndice C. Os arquivos AM e AF não foram incluídos na estatística. A Tabela 3.21 apresenta os limites superior e inferior para quantização do logaritmo natural do módulo de b_0 , obtidos estatisticamente e que incluem 99% das ocorrências do parâmetro.

Tabela 3.21 - Limites para quantização do coeficiente do filtro PLA.

Logaritmo Natural de $ b_0 $	
Máximo	0,0
Mínimo	-1,658525

A Fig.3.10 apresenta o histograma do logaritmo natural do módulo de b_0 , onde pode ser visto o acúmulo de ocorrências em $\ln|b_0| = 0$ (filtro estabilizado).

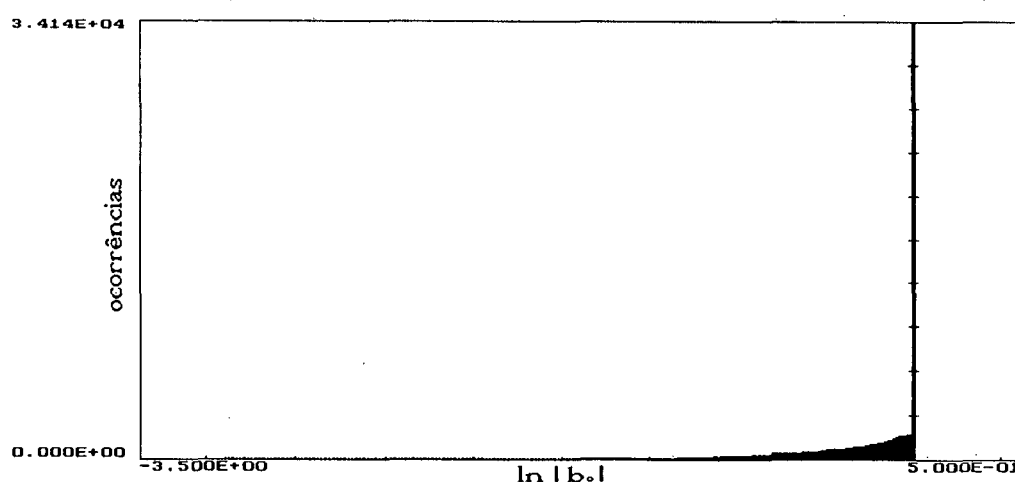


Fig.3.10 - Histograma do $\ln|b_0|$.

Para avaliar o nível de degradação introduzido pela limitação do parâmetro b_0 , foram simuladas quantizações com infinitos bits, utilizando os limites apresentados. A Tabela 3.22 apresenta os resultados objetivos obtidos, bem como aqueles obtidos com a não limitação de b_0 .

Tabela 3.22 - Resultados objetivos com a limitação de b_0 na quantização.

Arquivo	Procedimento	RSRSEG(dB)	Tp(min.)
AM/AF	b_0 não limitado	12,04/11,37	40/42
AM/AF	b_0 limitado	12,01/11,39	40/42

• Análise subjetiva informal: Não foram percebidas quaisquer degradações introduzidas pela limitação do coeficiente do filtro PLA.

A Tabela 3.23 apresenta os resultados objetivos obtidos em função da variação do número de bits na quantização de b_0 , utilizando-se os limites pré-estabelecidos.

Tabela 3.23 - Resultados objetivos em função do número de bits na quantização de b_0 .

Arquivo	Número de Bits *	RSRSEG(dB)	Tp(min.)
AM/AF	5	11,97/11,38	40/42
AM/AF	4	12,01/11,33	40/42
AM/AF	3	11,85/11,18	40/42
AM/AF	2	11,22/10,52	40/42

* incluindo 1 bit para o sinal de b_0 .

• Análise subjetiva informal: Foi constatado que a utilização de 4 ou 5 bits para quantização do logaritmo natural do módulo de b_0 , incluindo 1 bit para o sinal de b_0 , não introduz degradações perceptíveis. Já com a utilização de 2 ou 3 bits, pode ser notada uma pequena degradação, que entretanto não

compromete significativamente a qualidade final.

Para as simulações posteriores foi adotada a quantização do logaritmo natural do módulo de b_0 com 4 bits, incluindo 1 bit para o sinal de b_0 .

c) GANHO

Na quantização do ganho foi utilizado o procedimento de quantização uniforme do logaritmo natural do módulo deste parâmetro. Os limites superior e inferior foram determinados através da análise das ocorrências para os arquivos relacionados no Apêndice C. Os arquivos AM e AF não foram incluídos na estatística. A Tabela 3.24 apresenta os limites para o logaritmo natural do módulo do ganho obtidos estatisticamente e incluindo 99% das ocorrências deste parâmetro.

Tabela 3.24 - Limites para quantização de $\ln|g|$.

Logaritmo Natural do $ g $	
Máximo	4,753302
Mínimo	-0,142755

A influência da limitação do ganho foi avaliada através da quantização deste parâmetro com infinitos bits. Os resultados objetivos são apresentados na Tabela 3.25.

Tabela 3.25 - Resultados objetivos em função da limitação do ganho.

Arquivo	Procedimento	RSRSEG(dB)	Tp(min.)
AM/AF	S/limitação do ganho	12,01/11,33	40/42
AM/AF	C/limitação do ganho	11,98/11,34	42/45

● **Análise subjetiva informal:** Os resultados obtidos demonstraram que os limites adotados para o ganho não influenciam de forma perceptível a qualidade dos resultados.

O aumento do tempo de processamento no caso da limitação do ganho deve-se ao procedimento de quantização adotado no programa. A utilização de um procedimento de quantização através de tabelas deve reduzir bastante o acréscimo de tempo verificado.

A Tabela 3.26 apresenta os resultados objetivos obtidos em função da variação do número de bits utilizados para a quantização do logaritmo natural do módulo do ganho.

Tabela 3.26 - Resultados objetivos em função do número de bits na quantização do ganho.

Arquivo	Número de Bits *	RSRSEG(dB)	Tp(min.)
AM/AF	6	11,97/11,29	42/45
AM/AF	5	11,93/11,24	42/45
AM/AF	4	11,69/11,13	42/45
AM/AF	3	11,23/10,60	42/45
AM/AF	2	7,96/ 7,67	42/45

* incluindo 1 bit para o sinal do ganho

● **Análise subjetiva informal:** A utilização de 5 ou 6 bits na quantização do ganho não introduziu degradação sensível no sinal sintetizado. Com a utilização de 4 bits notou-se uma ligeira degradação, podendo ainda o resultado ser considerado de boa qualidade. Já com 3 bits, as degradações introduzidas foram mais significativas.

Para as simulações posteriores foi adotada a quantização do ganho com 5 bits, sendo 4 bits para representação do logaritmo natural do módulo do ganho e 1 bit para representação do sinal do ganho.

d) NÚMERO DE "CODEWORDS"

A Tabela 3.27 apresenta os resultados objetivos obtidos em função da variação do número de bits utilizado para codificar a "codeword" escolhida.

Tabela 3.27 - Resultados objetivos em função do número de "codewords".

Arquivo	"Codewords"(bits)	RSRSEG(dB)	Tp(min.)
AM/AF	1024(10)	11,93/11,24	42/45
AM/AF	512(9)	11,44/10,92	25/26
AM/AF	256(8)	11,12/10,57	16/17
AM/AF	128(7)	10,69/10,14	11/12
AM/AF	64(6)	10,19/ 9,69	9/10
AM/AF	32(5)	9,80/ 9,29	8/ 8
AM/AF	16(4)	9,27/ 8,66	7/ 8
AM/AF	8(3)	8,74/ 8,24	7/ 7
AM/AF	4(2)	8,00/ 7,63	7/ 7
AM/AF	2(1)	6,99/ 6,74	7/ 7
AM/AF	1(0)	6,11/ 5,48	7/ 7

● Análise subjetiva informal: A utilização de até 128 "codewords", no mínimo, proporcionou resultados de boa qualidade. A utilização de 1024 "codewords" apresentou um melhor resultado, tendo sido a diferença em relação à utilização de 512 "codewords" não muito significativa. Uma análise prévia demonstrou que a utilização de mais do que 1024 "codewords" não implica aumento perceptível na qualidade do sinal sintetizado, sendo portanto uma alternativa não recomendável devido ao conseqüente aumento da complexidade computacional envolvido.

Para as simulações posteriores foi mantido o "codebook" de 1024 "codewords".

e) TAXA FINAL

A adoção do número de bits especificado para a quantização de cada parâmetro em cada uma das simulações anteriormente realizadas, conduz a uma taxa final de 6800 bps. A taxa convencional de 7200 bps pode ser obtida com a diminuição do número de amostras dos quadros e/ou subquadros ou com o aumento do número de bits na quantização dos parâmetros.

Na fase de quantização, a etapa que resultou em maior redução na RSRSEG foi a quantização dos coeficientes do filtro PCA. Por este motivo, foi aumentado o número de bits na quantização destes parâmetros de maneira a obter-se a taxa final de 7200 bps.

A distribuição dos bits para a taxa de transmissão final de 7200 bps, utilizando-se quadros de 20 ms e subquadros de 5 ms, é apresentada na Tabela 3.28.

Tabela 3.28 - Distribuição de bits e taxas por etapa para o codificador CELP de 7200 bps.

Etapa	Número de Bits	Taxa(bps)
Filtro PCA	5+5+5+4+4+4+4+3+3+3	2000
Filtro PLA	7+3+1	2200
Ganho	4+1	1000
"Codebook"	10	2000
Total	-	7200

($N_q = 160$ amostras, $N = 40$ amostras)

A Tabela 3.29 apresenta os resultados objetivos obtidos com as taxas de 6800 bps e 7200 bps.

Tabela 3.29 - Desempenho objetivo do codificador CELP com taxas de 6800 bps e 7200 bps.

Arquivo	Taxa(bps)	RSRSEG(dB)	Tp(min.)
AM/AF	6800	11,93/11,24	42/45
AM/AF	7200	12,29/11,61	42/45

● Análise subjetiva informal: Pôde ser observada uma pequena melhoria na qualidade com taxa de 7200 bps em relação à taxa de 6800 bps.

Para as simulações posteriores foi mantida a configuração de bits adotada para o codificador CELP com taxa final de 7200 bps.

3.4.14 - DESEMPENHO EM TAXAS DIVERSAS

A análise do desempenho para várias taxas foi realizada fixando-se o número de bits na quantização de cada parâmetro e variando-se a duração dos quadros e subquadros. Os resultados objetivos são apresentados na Tabela 3.30.

Tabela 3.30 - Resultados objetivos com taxas entre 9600 bps e 2400 bps.

Arquivo	Quadro/Subquadro(amstras)	Taxa(bps)	RSRSEG(dB)	Tp(min.)
AM/AF	120/ 30	9600	13,84/13,11	43/46
AM/AF	160/ 40	7200	12,29/11,61	42/45
AM/AF	240/ 60	4800	9,97/ 9,57	42/44
AM/AF	320/ 80	3600	8,30/ 8,26	42/44
AM/AF	480/120	2400	6,63/ 6,57	42/44

● Análise subjetiva informal: Os resultados demonstraram excelente qualidade nas taxas de 9600 bps e 7200 bps; muito próximos daqueles obtidos com o modelo não quantizado. Já a 4800 bps, a qualidade pode ser considerada boa, apesar da ocorrência de maior degradação. Em 3600 bps os resultados foram

razoáveis, permitindo ainda a clara identificação do locutor e da locução. Em 2400 bps os resultados foram ruins, dificultando inclusive a identificação da locução pelo nível de ruído introduzido.

3.4.15 - VARIAÇÃO DO FATOR DE PONDERAÇÃO DO ERRO

A Tabela 3.31 apresenta os resultados objetivos obtidos com a variação do fator de ponderação do erro para o codificador CELP à taxa de 7200 bps.

Tabela 3.31 - Resultados objetivos em função da variação do fator de ponderação do erro em 7200 bps.

Arquivo	Fator de Ponderação	RSRSEG(dB)	Tp(min.)
AM/AF	0,80	11,69/11,09	42/45
AM/AF	0,82	11,81/11,19	42/45
AM/AF	0,84	11,96/11,42	42/45
AM/AF	0,86	12,06/11,50	42/45
AM/AF	0,88	12,22/11,61	42/45
AM/AF	0,90	12,29/11,61	42/45
AM/AF	0,92	12,36/11,61	42/45
AM/AF	0,94	12,39/11,63	42/45
AM/AF	0,96	12,39/11,66	42/45
AM/AF	0,98	12,28/11,71	42/45
AM/AF	1,00	12,08/11,53	42/45

● **Análise subjetiva informal:** Foi constatado que a utilização de fatores de ponderação do erro entre 0,8 e 0,9 conduziu a resultados muito semelhantes e com uma leve diminuição do ruído em relação aos resultados obtidos com fatores superiores.

Em vista dos resultados obtidos, foi mantido o fator de ponderação do erro igual a 0,9 para as simulações posteriores.

3.4.16 - VARIAÇÃO DO FATOR DE EXPANSÃO DA BANDA DO FILTRO PCA

A influência da variação do fator de expansão da banda do filtro PCA foi avaliada para a taxa de 7200 bps e os resultados objetivos são apresentados na Tabela 3.32.

Tabela 3.32 - Resultados objetivos em função do fator de expansão da banda do filtro PCA em 7200 bps.

Arquivo	Expansão(Hz)	Fator de Expansão	RSRSEG(dB)	Tp(min.)
AM/AF	50	0,980	12,64/11,78	42/45
AM/AF	30	0,988	12,55/11,74	42/45
AM/AF	15	0,994	12,29/11,61	42/45
AM/AF	10	0,996	12,25/11,60	42/45
AM/AF	5	0,998	12,14/11,52	42/45
AM/AF	0	1,000	12,01/11,44	42/45

● Análise subjetiva informal: Os resultados obtidos revelaram que a utilização de expansão da banda do filtro PCA entre 10 Hz e 30 Hz conduziu a uma melhoria na qualidade do sinal sintetizado.

Para as simulações posteriores, em vista dos resultados obtidos, foi mantido o fator de expansão da banda do filtro PCA em 0,994, o que equivale a uma expansão de banda em torno de 15 Hz.

3.4.17 - PÓS-FILTRAGEM

O desempenho do codificador CELP na taxa de 4800 bps com a utilização de pós-filtragem foi analisado e os resultados objetivos são apresentados na Tabela 3.33. Avaliações preliminares demonstraram que o uso de pós-filtragem nas taxas de 7200 bps e 9600 bps não implica melhoria sensível da qualidade do

sinal sintetizado, além do que aumenta a complexidade da etapa de síntese.

A Tabela 3.34 apresenta os resultados objetivos obtidos em 3600 bps e 2400 bps com a utilização de pós-filtragem.

Tabela 3.33 - Resultados objetivos em 4800 bps com pós-filtragem.

Arquivo	α	β	ϵ	RSRSEG(dB)	Tp(min.)
AM/AF	0,0	0,3	0,3	9,22/8,79	42/45
AM/AF	0,1	0,3	0,3	9,44/9,07	42/45
AM/AF	0,0	0,2	0,3	9,55/9,14	42/45
AM/AF	0,0	0,4	0,3	8,62/8,23	42/45
AM/AF	0,5	0,8	0,3	5,72/6,24	42/45
AM/AF	0,4	0,8	0,3	5,00/5,73	42/45
AM/AF	0,6	0,8	0,3	6,78/7,20	42/45
AM/AF	0,5	0,7	0,3	7,66/7,83	42/45
AM/AF	0,5	0,9	0,3	3,71/4,57	42/45
AM/AF	0,0	0,3	0,2	9,39/8,94	42/45
AM/AF	0,0	0,3	0,4	8,96/8,57	42/45
AM/AF	0,5	0,8	0,2	5,77/6,31	42/45
AM/AF	0,5	0,8	0,4	5,64/6,13	42/45

● Análise subjetiva informal: O efeito da pós-filtragem caracterizou-se, de uma forma geral, por um "abafamento" do sinal sintetizado, excessivo em alguns casos. Os melhores resultados foram obtidos com os parâmetros (0,6;0,8;0,3) e (0,0;0,2;0,3), tendo o primeiro caso implicado maior redução de agudos. O aumento subjetivo de qualidade foi evidente.

Tabela 3.34 - Resultados objetivos em 3600 bps e 2400 bps com utilização de pós-filtragem.

Arquivo	Taxa(bps)	α	β	ϵ	RSRSEG(dB)	Tp(min.)
AM/AF	3600	0,0	0,2	0,3	8,22/8,06	42/44
AM/AF	2400	0,0	0,2	0,3	6,67/6,57	42/44
AM/AF	3600	0,6	0,8	0,3	6,14/6,43	42/44
AM/AF	2400	0,6	0,8	0,3	5,10/5,46	42/44

• Análise subjetiva informal: Os resultados obtidos com os fatores (0,6;0,8;0,3) apresentaram maior eficiência na redução do ruído, implicando resultados de melhor qualidade.

Como conclusão, tem-se que a pós-filtragem é um procedimento eficaz na melhoria da qualidade do sinal sintetizado pelo codificador CELP em taxas inferiores a 7200 bps. A determinação dos fatores de pós-filtragem deve ser realizada com cuidado, a fim de ser encontrado o melhor resultado em termos subjetivos.

3.4.18 - VARIAÇÃO DO NÚMERO DE "CODEWORDS" PARA AS TAXAS BASE DE 7200 bps E 9600 bps

Como referência para comparações futuras quanto aos resultados objetivos e aos tempos de processamento, foram obtidos os resultados objetivos em função da redução do tamanho do "codebook", a partir das configurações de 9600 bps e 7200 bps com 1024 "codewords", apresentados nas Tabelas 3.35 e 3.36.

Tabela 3.35 - Resultados objetivos com a redução do número de "codewords" a partir da configuração básica de 9600 bps.

Arquivo	"Codewords"	Taxa(bps)	RSRSEG(dB)	Tp(min.)
AM/AF	1024	9600	13,84/13,11	43/46
AM/AF	512	9333	13,54/12,70	25/26
AM/AF	256	9067	13,09/12,40	16/17
AM/AF	128	8800	12,67/12,02	11/12
AM/AF	64	8533	12,17/11,52	9/10
AM/AF	32	8267	11,72/11,11	8/ 8
AM/AF	16	8000	11,14/10,62	7/ 8
AM/AF	8	7733	10,73/10,12	7/ 8
AM/AF	4	7467	9,97/ 9,50	7/ 7
AM/AF	2	7200	8,89/ 8,59	7/ 7
AM/AF	1	6933	8,31/ 8,08	7/ 7

Tabela 3.36 - Resultados objetivos com a redução do número de "codewords" a partir da configuração básica de 7200 bps.

Arquivo	"Codewords"	Taxa(bps)	RSRSEG(dB)	Tp(min.)
AM/AF	1024	7200	12,29/11,61	42/45
AM/AF	512	7000	11,88/11,25	25/26
AM/AF	256	6800	11,52/10,94	16/17
AM/AF	128	6600	11,16/10,51	11/12
AM/AF	64	6400	10,66/10,06	9/10
AM/AF	32	6200	10,22/ 9,67	8/ 8
AM/AF	16	6000	9,66/ 9,16	7/ 8
AM/AF	8	5800	9,06/ 8,53	7/ 7
AM/AF	4	5600	8,31/ 7,89	7/ 7
AM/AF	2	5400	7,43/ 7,13	7/ 7
AM/AF	1	5200	6,39/ 5,88	7/ 7

● **Análise subjetiva informal:** Tanto para 7200 bps como para 9600 bps, pôde ser observado um decréscimo na qualidade do sinal sintetizado com a redução do número de "codewords". Os resultados obtidos com no mínimo 128 "codewords" podem ser considerados bons.

3.5 - CONCLUSÕES

Neste capítulo foram estudados vários procedimentos para determinação e quantização escalar dos parâmetros do codificador CELP convencional. O desempenho das diversas configurações do codificador foi avaliado em termos objetivos (RSRSEG) e subjetivos (análise comparativa informal).

Na determinação dos parâmetros do filtro PCA, o método da autocorrelação com janela de Hamming mostrou-se o mais adequado. A utilização de 10 coeficientes, expansão da banda em torno de 15 Hz e a quantização escalar das razões log-área com um total de 40 bits conduziu a bons resultados.

Com relação ao filtro PLA, a utilização de um único coeficiente, com

módulo menor ou igual a 1 (filtro não instável) e quantização uniforme do logaritmo do módulo deste coeficiente com 4 bits, sendo 1 bit para o sinal, apresentou um bom compromisso entre qualidade e complexidade. O retardo M foi restrito à faixa de 33 a 160 e o procedimento de busca exaustiva do mesmo, juntamente com o procedimento de determinação do coeficiente do filtro em laço fechado, implicaram melhor qualidade do sinal sintetizado.

A quantização uniforme do logaritmo do módulo do ganho com 5 bits, sendo 1 bit para o sinal do ganho, apresentou resultados muito próximos daqueles obtidos com o modelo não quantizado.

Pôde ser observado o decréscimo de qualidade com a redução do número de "codewords" entretanto, utilizando-se um mínimo de 128 "codewords", os resultados puderam ser considerados bons.

A utilização dos procedimentos de ponderação do erro e pós-filtragem mostraram-se eficientes, aumentando a qualidade subjetiva do sinal sintetizado. O procedimento de pós-filtragem apresentou vantagens muito reduzidas em 7200 bps e 9600 bps, não justificando sua utilização nestas taxas.

Foram obtidas configurações operando entre 2400 bps e 9600 bps, todas derivadas da configuração de 7200 bps, variando-se as durações dos quadros e subquadros. Os resultados abaixo de 4800 bps apresentaram degradações mais acentuadas, enquanto que os resultados a 7200 bps e 9600 bps apresentaram muito boa qualidade. Excluído o problema da elevada complexidade computacional, os resultados obtidos mostraram que o codificador CELP convencional apresenta-se como uma ótima opção para redução da taxa de transmissão ou para o armazenamento econômico de sinais de voz com alta qualidade abaixo de 10 kbps.

CAPÍTULO 4

CODIFICAÇÃO CMSEC

4.1 - INTRODUÇÃO

Neste capítulo é apresentada uma estrutura de codificação paramétrica para a qual não foi encontrada qualquer referência anterior. A proposta está baseada no codificador CELP convencional e demonstrou ter possibilidade de obtenção de qualidade equivalente a este último, com menor complexidade computacional.

Uma parte significativa da complexidade computacional na codificação CELP convencional é devida à busca da melhor "codeword", já que cada candidata do "codebook" deve ser filtrada pelos filtros PLA e PCA para determinação do erro (análise por síntese). Na estrutura proposta, a excitação é formada pela combinação linear de várias seqüências randômicas obtidas de "codebooks" de tamanho reduzido. Tal estrutura será denominada CMSEC ("Constrained Multiple-Stage Excitation Coding").

4.2 - DETERMINAÇÃO DOS PARÂMETROS DO CODIFICADOR CMSEC

A Fig.4.1 e a Fig.4.2 apresentam respectivamente os modelos de análise e de síntese do codificador CMSEC.

A única diferença no modelo de análise do codificador CMSEC em relação ao CELP convencional é quanto à determinação da excitação. Neste caso, são

utilizados L "codebooks", cada um com um número reduzido de "codewords". A excitação será formada pela combinação linear de L seqüências randômicas, cada uma ponderada por um ganho g_l , que minimiza o erro ponderado $\hat{e}(n)$. A busca do melhor conjunto de "codewords" é realizada, como no modelo CELP convencional, de forma exaustiva, porém, para um número reduzido de "codewords". O número efetivo de seqüências a serem filtradas pela cascata PLA/PCA, para efeito de complexidade, corresponde a L vezes o número de "codewords" em cada "codebook". A possibilidade de conjuntos de "codewords" diferentes fica restrita pelo fato de ser utilizado somente um índice, que simultaneamente referencia uma "codeword" de cada "codebook".

Os codificadores MSVXC ("Multiple-Stage Vector Excitation Coding") [56] e VSELP ("Vector Sum Excited Linear Predictor") [2,55] têm estruturas semelhantes ao CMSEC. Nestes, entretanto, são enviados índices distintos para cada "codebook" e os ganhos não são determinados simultaneamente.

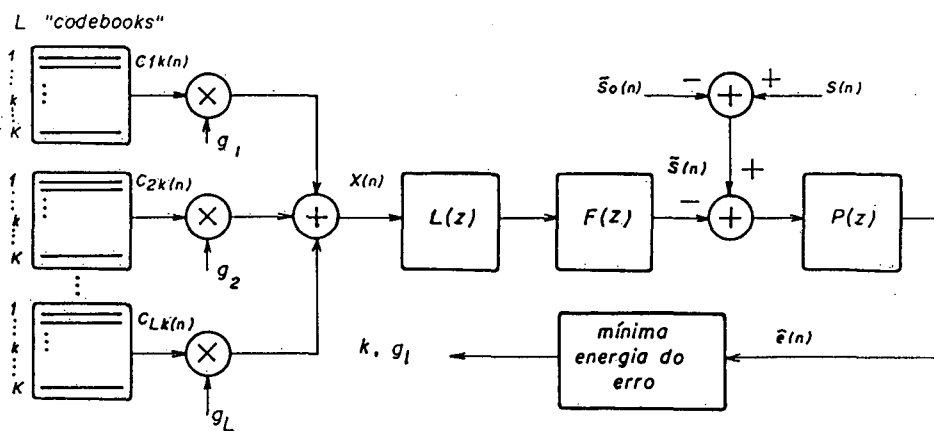


Fig.4.1 - Modelo de análise do codificador CMSEC.

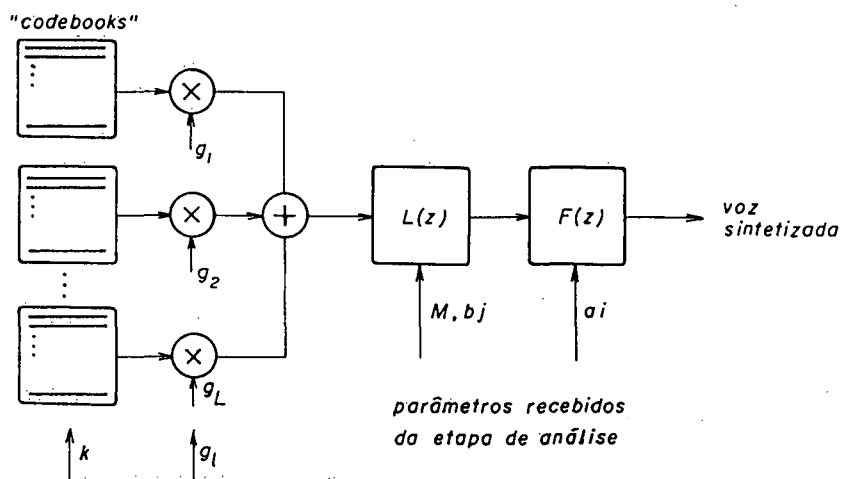


Fig.4.2 - Modelo de síntese do codificador CMSEC.

Seja $x(n)$ a excitação dos filtros PLA e PCA dada por:

$$x(n) = \sum_{l=1}^L g_l \cdot c_{lk}(n) \quad (4.1)$$

onde:

g_l são os múltiplos ganhos;

$c_{lk}(n)$ é k -ésima "codeword" do "codebook" l ;

Como o sistema é linear na etapa de determinação da excitação (filtros com memória inicial nula), a saída do filtro combinado PLA/PCA é dada por:

$$\tilde{s}(n) = \sum_{l=1}^L g_l \cdot c_{lk}(n) * l(n) * f(n) \quad (4.2)$$

onde:

$l(n)$ é a resposta ao impulso do filtro PLA;

$f(n)$ é a resposta ao impulso do filtro PCA.

O erro ponderado é então:

$$\hat{e}(n) = [s(n) - \tilde{s}_0(n)] * p(n) + p_0(n) - \sum_{l=1}^L g_l \cdot c_{lk}(n) * l(n) * f(n) * p(n) \quad (4.3)$$

É desejado minimizar a energia E do erro ponderado $\hat{e}(n)$ dentro de um intervalo de amostras correspondente a um subquadro. Seja a energia do erro ponderado dada por:

$$E = \sum_{n=0}^{N-1} [\hat{e}^2(n)] \quad (4.4)$$

onde N é o número de amostras do subquadro.

Os valores de g_l que minimizam E são encontrados igualando-se a zero as derivadas parciais de E com relação a cada um dos ganhos g_l .

$$\begin{aligned} \frac{\partial E}{\partial g_q} &= 2 \cdot \sum_{n=0}^{N-1} \{ [s(n) - \tilde{s}_0(n)] * p(n) + p_0(n) - \sum_{l=1}^L g_l \cdot c_{lk}(n) * l(n) * f(n) * p(n) \} \cdot \\ &\quad \cdot [c_{qk}(n) * l(n) * f(n) * p(n)] = 0, \quad 1 \leq q \leq L \end{aligned} \quad (4.5)$$

O procedimento descrito resulta no sistema linear de L equações e L incógnitas dado a seguir:

$$\begin{aligned} &\sum_{l=1}^L g_l \cdot \sum_{n=0}^{N-1} \{ [c_{lk}(n) * l(n) * f(n) * p(n)] \cdot [c_{qk}(n) * l(n) * f(n) * p(n)] \} = \\ &= \sum_{n=0}^{N-1} \{ [s(n) - \tilde{s}_0(n)] * p(n) + p_0(n) \} \cdot [c_{qk}(n) * l(n) * f(n) * p(n)] \quad , 1 \leq q \leq L \end{aligned} \quad (4.6)$$

A matriz obtida do lado esquerdo de (4.6) é simétrica, o que permite a resolução do sistema através da decomposição de Cholesky.

Pode ser observado que, no caso de utilizar-se somente um ganho ($L=1$), as equações representadas por (4.6) reduzem-se exatamente à equação para determinação do ganho no modelo CELP convencional.

4.3 - QUANTIZAÇÃO DOS PARÂMETROS DO CODIFICADOR CMSEC

De forma análoga ao codificador CELP convencional, os ganhos são os últimos parâmetros a serem quantizados e sua determinação pode ser feita com base nos demais parâmetros já quantizados. Por outro lado, a utilização de vários ganhos e o tamanho reduzido dos "codebooks" tornam mais crítica a quantização dos mesmos.

A quantização dos ganhos pode ser realizada em conjunto (quantização vetorial) ou isoladamente (quantização escalar). É esperado que os procedimentos de quantização vetorial [90], não abordados neste trabalho, devam conduzir a uma menor quantidade de bits para quantização dos parâmetros ao custo de um aumento na complexidade.

Um procedimento iterativo de otimização dos ganhos na quantização escalar consiste em quantizá-los um a um e, para cada ganho quantizado, recalculando os demais de forma a minimizar a energia do erro ponderado. Assim, são determinados os L ganhos não quantizados através de (4.6). O ganho g_1 é então quantizado e seu valor é utilizado para reduzir a ordem do sistema de equações, gerando um sistema de $L-1$ equações e incógnitas. Após recalculados os ganhos $g_2 \dots g_L$, o ganho g_2 é quantizado e utilizado na redução da ordem do sistema. O procedimento é executado até que o último ganho g_L seja determinado e quantizado. Nota-se que a cada iteração somente o vetor correspondente aos

termos independentes de (4.6) é recalculado, permanecendo inalterados os elementos da matriz inicial dos coeficientes e portanto conservando a propriedade de simetria.

4.4 - ANÁLISE DO DESEMPENHO DO CODIFICADOR CMSEC

Com o objetivo de verificar o desempenho do codificador CMSEC, realizaram-se várias simulações onde foram variados o número de ganhos ("codebooks"), o número de bits na quantização de cada ganho e o número de "codewords", tendo sido os demais parâmetros mantidos de maneira idêntica aqueles utilizados no codificador CELP de 7200 bps anteriormente analisado.

A Tabela 4.1 apresenta os resultados objetivos obtidos para diversas configurações do codificador CMSEC. O número de ganhos e de "codewords" foi escolhido de forma que todas as configurações apresentem possibilidades de taxas finais entre 7200 e 9600 bps, quando utilizada a quantização escalar dos ganhos com 3, 4 ou 5 bits. Os resultados objetivos apresentados foram obtidos com quantização infinita dos ganhos.

Ainda com relação aos resultados objetivos, algumas configurações do codificador CMSEC apresentaram possibilidades de RSRSEG maiores do que as do codificador CELP convencional com possibilidade de operação à mesma taxa. É importante ressaltar que neste ponto as comparações estão sendo realizadas com os resultados obtidos com infinitos bits na quantização dos ganhos e, na prática, a confirmação destes resultados depende do procedimento de quantização dos ganhos adotado.

Tabela 4.1 - Resultados objetivos para o codificador CMSEC com quantização infinita dos ganhos.

Arquivo Ganhos Codewords			Taxa(bps)			RSRSEG(dB)	Tp(min.)
			3bit/g	4bit/g	5bit/g	(∞ bit/g)	
AM/AF	1	1024	6800	7000	7200	12,39/11,69	47/50
AM/AF	2	128	6800	7200	7600	12,32/11,62	16/17
AM/AF	2	256	7000	7400	7800	12,77/11,97	25/26
AM/AF	3	16	6800	7400	8000	11,83/11,10	8/ 9
AM/AF	3	32	7000	7600	8200	12,39/11,54	10/11
AM/AF	3	64	7200	7800	8400	12,83/11,98	13/14
AM/AF	3	128	7400	8000	8600	13,22/12,33	20/21
AM/AF	3	256	7600	8200	8800	13,58/12,77	33/35
AM/AF	4	8	7200	8000	8800	12,28/11,37	8/ 9
AM/AF	4	16	7400	8200	9000	12,76/11,80	9/10
AM/AF	4	32	7600	8400	9200	13,15/12,28	11/12
AM/AF	4	64	7800	8600	9400	13,67/12,68	16/17
AM/AF	4	128	8000	8800	9600	14,05/13,12	24/26
AM/AF	5	2	7400	8400	9400	11,57/10,77	7/ 8
AM/AF	5	4	7600	8600	9600	12,53/11,59	8/ 8
AM/AF	5	8	7800	8800	9800	13,07/12,08	8/ 9
AM/AF	5	16	8000	9000	10000	13,48/12,53	10/10
AM/AF	5	32	8200	9200	10200	13,88/12,92	12/13
AM/AF	5	64	8400	9400	10400	14,35/13,34	18/19
AM/AF	5	128	8600	9600	10600	14,85/13,75	30/31
AM/AF	6	2	8000	9200	10400	12,41/11,51	7/ 8
AM/AF	6	4	8200	9400	10600	13,34/12,25	8/ 8
AM/AF	6	8	8400	9600	10800	13,64/12,59	9/ 9
AM/AF	6	16	8600	9800	11000	14,14/13,09	10/11
AM/AF	6	32	8800	10000	11200	14,52/13,53	14/15
AM/AF	6	64	9000	10200	11400	14,94/13,99	21/22
AM/AF	6	128	9200	10400	11600	15,38/14,35	35/37
AM/AF	7	1	8400	9800	11200	12,08/11,19	7/ 8
AM/AF	7	2	8600	10000	11400	12,98/12,01	7/ 8
AM/AF	7	4	8800	10200	11600	13,53/12,66	8/ 8
AM/AF	7	8	9000	10400	11800	14,15/13,16	9/ 9
AM/AF	7	16	9200	10600	12000	14,69/13,67	11/12
AM/AF	7	32	9400	10800	12200	15,19/14,15	15/16
AM/AF	7	64	9600	11000	12400	15,57/14,55	24/25
AM/AF	8	1	9000	10600	12200	12,68/11,73	7/ 8
AM/AF	8	2	9200	10800	12400	13,63/12,63	7/ 8
AM/AF	8	4	9400	11000	12600	14,34/13,28	8/ 9
AM/AF	8	8	9600	11200	12800	14,85/13,67	9/10
AM/AF	8	16	9800	11400	13000	15,27/14,22	12/12
AM/AF	8	32	10000	11600	13200	15,79/14,73	17/18
AM/AF	8	64	10200	11800	13400	16,12/15,14	27/28
AM/AF	9	1	9600	11400	13200	12,78/12,12	7/ 8

● Análise subjetiva informal: Todas as simulações apresentaram bons resultados, de uma forma geral. As simulações onde foram utilizadas 1 ou 2 "codewords" em cada "codebook", apresentaram-se ligeiramente mais ruidosas, fato mais evidente na audição com fones de ouvido.

Dos resultados obtidos, foram escolhidas quatro configurações para análise do desempenho frente à quantização dos ganhos e que são relacionadas a seguir:

- a) melhor resultado objetivo com possibilidade de operação a 7200 bps: 3 ganhos e 64 "codewords";
- b) menor taxa possível com resultado objetivo comparável ao do codificador CELP a 7200 bps: 2 ganhos e 256 "codewords";
- c) melhor resultado objetivo com o menor tempo de processamento: 8 ganhos e 2 "codewords" (a alternativa com 9 ganhos e 1 "codeword" apresentou ruído mais acentuado);
- d) melhor resultado objetivo com possibilidade de operação a 9600 bps e 4 bits/ganho: 5 ganhos e 128 "codewords".

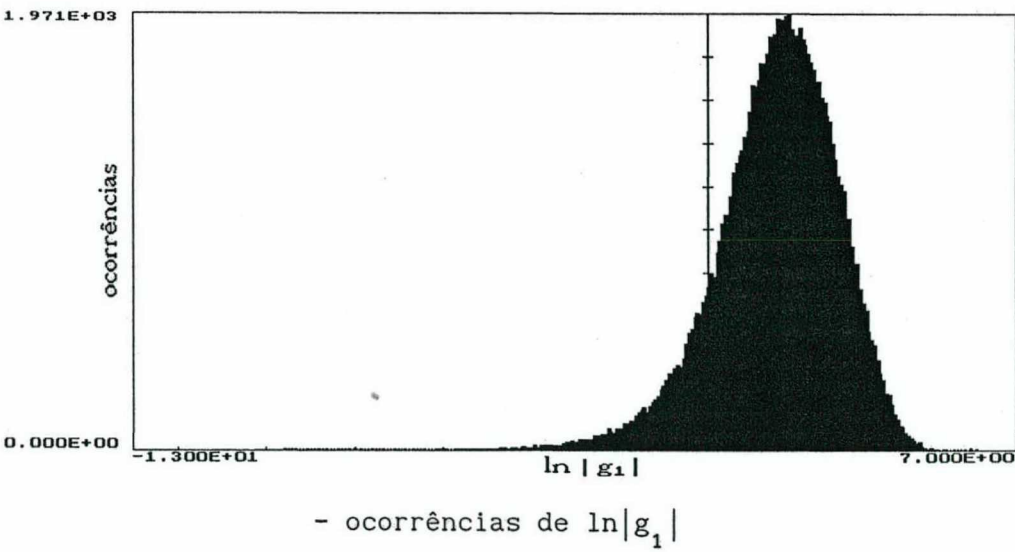
Para as quatro configurações escolhidas foram determinados os limites de quantização de cada ganho através da análise das ocorrências para os arquivos mencionados no apêndice C. Os limites foram determinados de forma a incluir 99% das ocorrências do logaritmo natural do módulo de cada ganho. A Tabela 4.2 apresenta os valores obtidos.

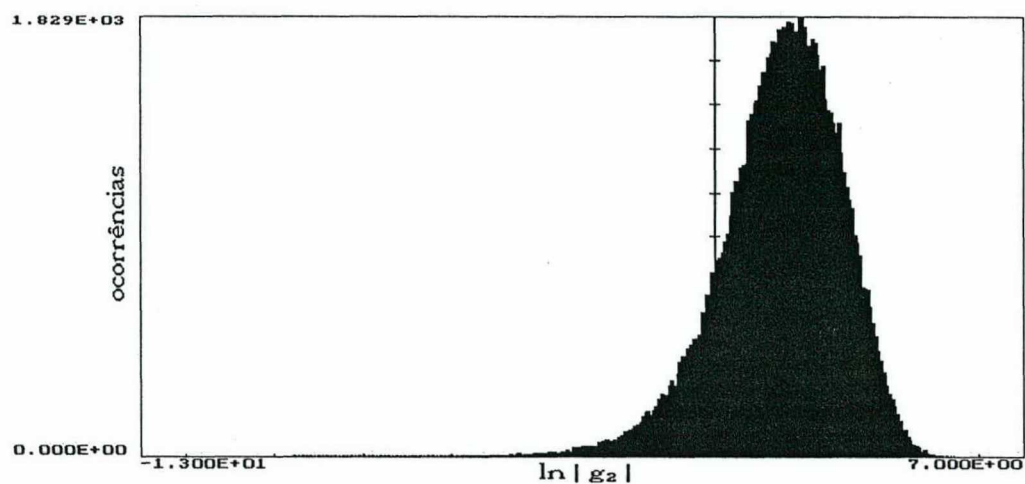
Ainda com relação às estatísticas dos ganhos, o levantamento das ocorrências considerando os ganhos anteriores já quantizados no procedimento de quantização iterativa, deve conduzir teoricamente a um melhor resultado do que aquele obtido com os ganhos anteriores não quantizados. Devido à quantidade de processamento envolvido, optou-se pela segunda alternativa, tendo entretanto uma análise informal revelado muito pouca diferença nos

limites dos ganhos obtidos com uma ou outra alternativa. A Fig.4.3 apresenta os histogramas com as ocorrências do logaritmo natural do módulo dos ganhos do codificador CMSEC para a configuração "a".

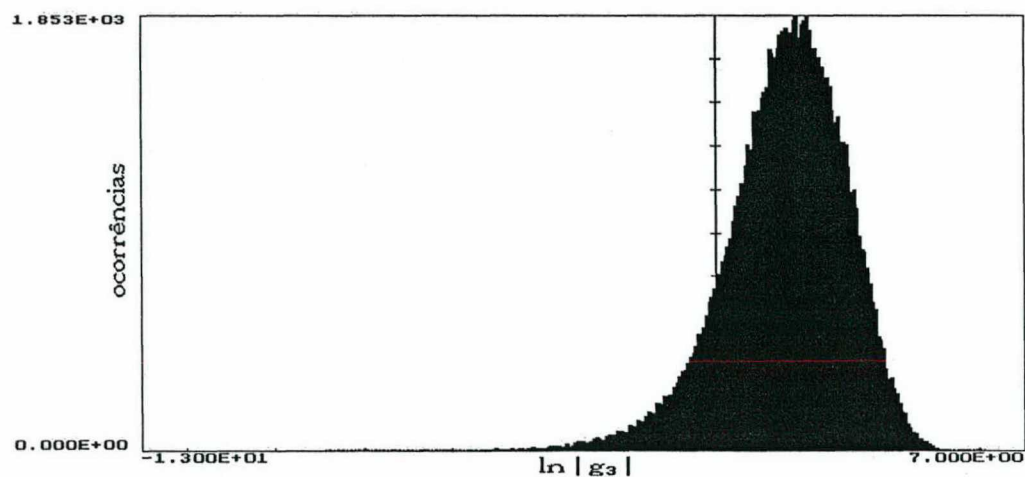
Tabela 4.2 - Limites para quantização dos ganhos no codificador CMSEC obtidos estatisticamente.

Config.		Limite $\ln g_1 $	$\ln g_2 $	$\ln g_3 $	$\ln g_4 $	$\ln g_5 $	$\ln g_6 $	$\ln g_7 $	$\ln g_8 $
a	máx.	4,433	4,373	4,399	-	-	-	-	-
	mín.	-3,154	-3,253	-3,163	-	-	-	-	-
b	máx.	4,561	4,552	-	-	-	-	-	-
	mín.	-2,750	-2,885	-	-	-	-	-	-
c	máx.	3,977	4,399	3,891	3,780	4,196	3,936	4,298	4,129
	mín.	-3,809	-3,360	-3,659	-3,834	-3,385	-3,786	-3,522	-3,799
d	máx.	4,334	4,382	4,445	4,312	4,321	-	-	-
	mín.	-3,284	-3,244	-3,430	-3,436	-3,277	-	-	-





- ocorrências de $\ln|g_2|$



- ocorrências de $\ln|g_3|$

Fig.4.3 - Ocorrências de $\ln|g_1|$ no codificador CMSEC de 3 ganhos.

As Tabelas 4.3 a 4.6 apresentam os resultados objetivos obtidos para as quatro configurações do codificador CMSEC, com limitação dos ganhos e quantização do logaritmo natural do módulo destes parâmetros com diversos números de bits e otimização iterativa.

Tabela 4.3 - Resultados objetivos obtidos com o codificador CMSEC de 3 ganhos e 64 "codewords".

Arquivo	Taxa(bps)			RSRSEG(dB)	Tp(min.)
	3bit/g	4bit/g	5bit/g		
AM/AF	∞	∞	∞	12,70/11,91	13/14
AM/AF	-	-	8400	12,47/11,75	13/14
AM/AF	-	7800	-	12,08/11,31	13/14
AM/AF	7200	-	-	10,87/10,35	13/14

● Análise subjetiva informal: Para o codificador CMSEC de 8400 bps, o resultado obtido foi equivalente ao do CELP convencional de 7200 bps. Para as taxas de 7800 e 7200 bps, foram observadas degradações, que embora não muito significativas, tornaram ambos os resultados ligeiramente inferiores ao CELP de 7200 bps.

Tabela 4.4 - Resultados objetivos obtidos com o codificador CMSEC de 2 ganhos e 256 "codewords".

Arquivo	Taxa(bps)			RSRSEG(dB)	Tp(min.)
	3bit/g	4bit/g	5bit/g		
AM/AF	∞	∞	∞	12,67/12,01	25/26
AM/AF	-	-	7800	12,40/11,79	25/26
AM/AF	-	7400	-	12,05/11,44	25/26
AM/AF	7000	-	-	11,03/10,59	25/26

● Análise subjetiva informal: O resultado obtido com o codificador CMSEC de 2 ganhos e 256 "codewords" a 7800 bps foi praticamente idêntico aquele obtido com o CELP convencional a 7200 bps. Já a 7400 bps e a 7000 bps foi possível observar degradações, mais acentuadas na última, não podendo contudo os resultados ser considerados ruins.

Tabela 4.5 - Resultados objetivos obtidos com o codificador CMSEC de 8 ganhos e 2 "codewords".

Arquivo	Taxa(bps)			RSRSEG(dB)	Tp(min.)
	3bit/g	4bit/g	5bit/g		
AM/AF	∞	∞	∞	13,54/12,49	7/ 8
AM/AF	-	-	12400	13,05/12,11	7/ 8
AM/AF	-	10800	-	11,88/11,12	7/ 8
AM/AF	9200	-	-	7,91/ 8,03	7/ 8

● Análise subjetiva informal: O codificador CMSEC de 8 ganhos e 2 "codewords", operando a 12400 bps, apresentou um resultado ligeiramente mais ruidoso do que o CELP a 7200 bps, tendo sido, ainda assim, um resultado de boa qualidade. A 10800 bps pode ser notada uma maior degradação e a 9600 bps o resultado foi ruim.

Tabela 4.6 - Resultados objetivos obtidos com o codificador CMSEC de 5 ganhos e 128 "codewords".

Arquivo	Taxa(bps)			RSRSEG(dB)	Tp(min.)
	3bit/g	4bit/g	5bit/g		
AM/AF	∞	∞	∞	14,78/13,79	30/31
AM/AF	-	-	10600	14,18/13,39	30/31
AM/AF	-	9600	-	13,58/12,80	30/31
AM/AF	8600	-	-	12,00/11,49	30/31

● Análise subjetiva informal: Os resultados obtidos com o codificador CMSEC de 5 ganhos e 128 "codewords" nas taxas de 10600 e 9600 bps foram praticamente idênticos àquele obtido com o codificador CELP a 9600 bps. Já a 8600 bps, o resultado é ligeiramente inferior.

De maneira geral, os resultados obtidos com o codificador CMSEC

apresentaram-se menos "abafados" do que aqueles obtidos com o CELP convencional, o que torna um pouco mais evidente o ruído de fundo. Nos casos onde o ruído for mais perceptível, um ajuste do fator de ponderação do erro pode ser uma solução adequada.

Resultado particularmente interessante foi obtido a 9600 bps, utilizando-se 128 "codewords" e 5 ganhos, todos com a mesma limitação imposta ao ganho do codificador CELP convencional ($-0,143 \leq \ln|g| \leq 4,753$). Neste caso, os resultados objetivos para os arquivos AM e AF (13,94 dB/13,12 dB) superaram aqueles obtidos com o CELP a 9600 bps, com menor complexidade. A qualidade subjetiva foi praticamente a mesma e até ligeiramente superior para o arquivo AF. Tal resultado conduz à conclusão de que deve ser estudado um procedimento mais adequado para quantização dos ganhos no codificador CMSEC, que implique em aumento da qualidade do sinal sintetizado sem aumento da taxa final em relação ao procedimento anteriormente adotado.

A Tabela 4.7 apresenta os resultados objetivos obtidos com taxas de 4800 bps, com e sem pós-filtragem, com os mesmos limites utilizados nas simulações da Tabela 4.6. A taxa de 4800 bps foi obtida a partir do aumento do número de amostras dos quadros e subquadros.

Tabela 4.7 - Resultados objetivos do codificador CMSEC a 4800 bps.

Arquivo	Ganhos	Codewords	RSRSEG(dB)	Tp(min.)	Condição
AM/AF	5	128	9,32/9,09	28/30	s/pós-filt.
AM/AF	5	128	9,10/8,82	28/30	c/pós-filt.

- **Análise subjetiva informal:** Sem pós-filtragem, os resultados obtidos têm uma característica "metálica", devido à utilização de um quadro maior do que no codificador CELP de mesma taxa. A utilização de pós-filtragem atenuou

significativamente o problema, sendo contudo o resultado final ainda ligeiramente inferior ao do CELP convencional.

4.5 - COMPLEXIDADE COMPUTACIONAL DOS CODIFICADORES CELP E CMSEC

A seguir são relacionadas expressões para quantificação do número de operações aritméticas na determinação da excitação nos codificadores CELP convencional e CMSEC:

a) Codificador CELP:

- Filtragem com PLA:

multiplicações: $K.T.(N-M+1)$, $N \geq M$

0 , $N < M$

adições: $K.T.(N-M+1)$, $N \geq M$

0 , $N < M$

- Filtragem com PCA:

multiplicações: $K.OP.[N-(OP+1)/2]$

adições: $K.OP.[N-(OP+1)/2]$

- Cálculo do ganho:

multiplicações: $K.(2.N+1)$

adições: $2.K.(N-1)$

- Amplificação da saída do PCA:

multiplicações: $K.N$

adições: -

- Cálculo do erro:

multiplicações: $K.N$

adições: $K.(2.N-1)$

b) Codificador CMSEC:

- Filtragem com PLA:

multiplicações: $L.K.T.(N-M+1)$, $N \geq M$

0 , $N < M$

adições: $L.K.T.(N-M+1)$, $N \geq M$

0 , $N < M$

- Filtragem com PCA:

multiplicações: $L.K.OP.[N-(OP+1)/2]$

adições: $L.K.OP[N-(OP+1)/2]$

- Montagem do sistema:

multiplicações: $L.K.N.[2+(L-1)/2]$

adições: $L.K.(N-1).[2+(L-1)/2]$

- Resolução do sistema:

multiplicações: $L.K.[1+5.(L-1)/2+(L-1).(L-2)/3]$, $L > 1$

adições: $K.[3.L.(L-1)/2+L.(L-1).(L-2)/6-1]$, $L > 1$

- Otimização dos ganhos:

multiplicações: $K.\{1+L.(L-1)/2+\sum_{i=2}^{L-1} i.[1+5.(i-1)/2+(i-1).(i-2)/3]\}$

, $L > 2$

$2.K$, $L = 2$

adições: $K.\{L.(L-1)/2+\sum_{i=2}^{L-1} [3.i.(i-1)/2+ i.(i-1).(i-2)/6-1]\}$, $L > 2$

K , $L = 2$

- Combinação das "codewords":

multiplicações: $L.K.N$

adições: $K.(L-1).N$

- Cálculo do erro:

multiplicações: $K.N$

adições: $K.(2.N-1)$

As Tabelas 4.8 a 4.12 apresentam o número de operações aritméticas calculadas através das expressões anteriores para os codificadores CELP de 7200 bps e 9600 bps e para os codificadores CMSEC analisados no item 4.4. Como o número de operações aumenta com a diminuição de M , e este pode variar entre 33 e 160 amostras, dependendo do sinal analisado, foi considerado o pior caso, $M = 33$. Pode ser verificada a menor complexidade do codificador CMSEC em todas as configurações apresentadas. Para o codificador CMSEC de 8 ganhos e 2 "codewords", verificou-se uma redução de operações aritméticas da ordem de 50 vezes em relação às configurações do CELP convencional apresentadas.

Tabela 4.8 - Operações aritméticas para determinação da excitação no codificador CELP convencional operando a 7200 bps e 9600 bps.

Procedimento	Multiplicações		Adições	
	7200 bps	9600 bps	7200 bps	9600 bps
filtragem com PLA	8192	-	8192	-
filtragem com PCA	353280	250880	353280	250880
cálculo do ganho	82944	62464	79872	59392
amplificação	40960	30720	-	-
cálculo do erro	40960	30720	80896	60416
total/subquadro	526336	374784	522240	370688
total/segundo	105267200	99942400	104448000	98850133
K = 1024, OP = 10, T = 1, M = 33, N = 40 (7200) / 30 (9600)				

Tabela 4.9 - Operações aritméticas para determinação da excitação no codificador CMSEC de 5 ganhos e 128 "codewords".

Procedimento	Multiplicações	Adições
filtragem com PLA	5120	5120
filtragem com PCA	220800	220800
montagem do sistema	102400	99840
solução do sistema	9600	4992
otimização dos ganhos	10240	5376
combinação das "codewords"	25600	20480
cálculo do erro	5120	10112
total/subquadro	378880	366720
total/segundo	75776000	73344000
K = 128, OP = 10, T = 1, M = 33, L = 5, N = 40		

Tabela 4.10 - Operações aritméticas para determinação da excitação no codificador CMSEC de 2 ganhos e 256 "codewords".

Procedimento	Multiplicações	Adições
filtragem com PLA	4096	4096
filtragem com PCA	176640	176640
montagem do sistema	51200	49920
solução do sistema	1792	512
otimização dos ganhos	512	256
combinação das "codewords"	20480	10240
cálculo do erro	10240	20224
total/subquadro	264960	261888
total/segundo	52992000	52377600
K = 256, OP = 10, T = 1, M = 33, L = 2, N = 40		

Tabela 4.11 - Operações aritméticas para determinação da excitação no codificador CMSEC de 3 ganhos e 64 "codewords".

Procedimento	Multiplicações	Adições
filtragem com PLA	1536	1536
filtragem com PCA	66240	66240
montagem do sistema	23040	22464
solução do sistema	1280	576
otimização dos ganhos	704	320
combinação das "codewords"	7680	5120
cálculo do erro	2560	5056
total/subquadro	103040	101312
total/segundo	20608000	20262400
K = 64, OP = 10, T = 1, M = 33, L = 3, N = 40		

Tabela 4.12 - Operações aritméticas para determinação da excitação no codificador CMSEC de 8 ganhos e 2 "codewords".

Procedimento	Multiplicações	Adições
filtragem com PLA	128	128
filtragem com PCA	5520	5520
montagem do sistema	3520	3432
solução do sistema	520	278
otimização dos ganhos	952	520
combinação das "codewords"	640	560
cálculo do erro	80	158
total/subquadro	11360	10596
total/segundo	2272000	2119200
K = 2, OP = 10, T = 1, M = 33, L = 8, N = 40		

4.6 - CONCLUSÕES

Os resultados obtidos com quantização infinita dos ganhos demonstraram que o codificador CMSEC tem possibilidade de produzir voz sintetizada com qualidade comparável àquela obtida com o codificador CELP convencional, operando abaixo de 10 kbps.

Através da escolha adequada do número de ganhos ("codebooks"), do número de "codewords" em cada "codebook" e do número de bits para quantização dos ganhos é possível obter um compromisso entre taxa, qualidade do sinal sintetizado e complexidade computacional que satisfaça os requisitos procurados.

A utilização de quantização escalar dos ganhos do codificador CMSEC conduziu a resultados de boa qualidade, com taxas finais ligeiramente superiores às obtidas com o codificador CELP convencional, com qualidade equivalente dos sinais de voz sintetizados com um e outro procedimento. Apesar do acréscimo na taxa, a complexidade computacional no CMSEC foi reduzida em relação ao CELP, em alguns casos significativamente.

Subjetivamente, os resultados obtidos com o codificador CMSEC apresentaram-se um pouco menos "abafados" do que aqueles obtidos com o CELP, sendo por este motivo o ruído de fundo ligeiramente mais acentuado. Nos casos onde o nível de ruído de fundo for julgado inconveniente ("codebooks" muito reduzidos), o ajuste do fator de ponderação do erro ou ainda a utilização de pós-filtragem podem ser soluções adequadas.

Pôde também ser percebida a necessidade do estudo de um procedimento de quantização mais eficiente para os ganhos do codificador CMSEC, que não implique em aumento considerável da taxa final para manutenção da qualidade em relação ao codificador CELP. Procedimentos de quantização vetorial ou de quantização escalar considerando a distribuição estatística dos parâmetros podem ser soluções adequadas.

CAPÍTULO 5

QUANTIZAÇÃO ESTATÍSTICA DO GANHO5.1 - INTRODUÇÃO

O procedimento de quantização uniforme do logaritmo do módulo do ganho com 5 bits apresenta resultados satisfatórios no codificador CELP convencional. Já no codificador CMSEC, a necessidade de utilização de menos bits na quantização de cada ganho, visando a não elevação da taxa final, implicou em resultados menos satisfatórios em alguns casos. Conseqüentemente, o estudo de um procedimento alternativo de quantização escalar dos parâmetros ganho faz-se necessário, tendo como objetivo o aumento da qualidade do sinal sintetizado para uma mesma taxa de transmissão, ou a diminuição desta taxa sem deterioração significativa da qualidade do sinal de voz sintetizado.

Neste capítulo, é proposto um método para a quantização escalar dos ganhos, que leva em consideração a função densidade de probabilidade destes parâmetros na distribuição dos níveis de quantização. O procedimento é aplicado aos codificadores CELP e CMSEC e seu desempenho é comparado com o procedimento de quantização uniforme do logaritmo do módulo do ganho, em termos da qualidade objetiva e subjetiva do sinal de voz sintetizado [103].

5.2 - PROCEDIMENTO DE QUANTIZAÇÃO ESTATÍSTICA DO GANHO

O levantamento das ocorrências do logaritmo do módulo do ganho, tanto no codificador CELP convencional quanto no codificador CMSEC, apresentou uma função densidade de probabilidade aproximadamente normal. A Fig.5.1 apresenta o histograma das ocorrências do logaritmo natural do módulo do ganho no codificador CELP convencional, superposto à função densidade de probabilidade normal com média e desvio padrão obtidos a partir dos dados levantados. Os valores foram obtidos através do processamento dos arquivos com segmentos de voz relacionados no Apêndice C. Os parâmetros utilizados no codificador CELP foram os mesmos definidos para a taxa de 7200 bps e especificados no item 3.4, tendo como única diferença a não quantização do parâmetro ganho.

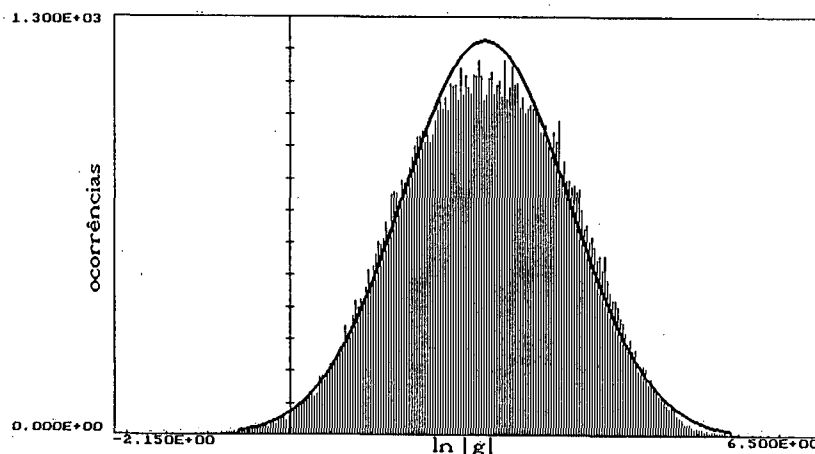


Fig.5.1 - Histograma do $\ln|g|$ no codificador CELP e densidade de probabilidade normal com mesmas média e variância.

Tendo o histograma do logaritmo do módulo do ganho no processo de

codificação CELP apresentado uma distribuição próxima da distribuição normal, a alocação dos níveis de quantização com base nesta mesma distribuição, ou seja, concentrando mais níveis nas zonas de maior ocorrência, pode conduzir a uma melhor qualidade do sinal sintetizado ou à necessidade de menos bits para manutenção da qualidade do sinal de voz sintetizado.

O procedimento adotado para a quantização estatística do ganho consiste em transformar adequadamente os valores do logaritmo do módulo do ganho em um parâmetro com probabilidade de ocorrência uniforme e limitado entre 0 e 1, e então quantizar uniformemente este novo parâmetro. Os valores quantizados sofrem uma transformação inversa, ou seja, de um parâmetro com distribuição uniforme para um parâmetro com distribuição normal.

Dado um parâmetro x com função densidade de probabilidade $f(x)$, a função $F(x)$ que transforma x em um parâmetro y com função densidade de probabilidade constante e limitado entre 0 e 1, é dada pelo teorema fundamental da função densidade [104]. Para uma distribuição normal, as funções $f(x)$ e $F(x)$ correspondem, respectivamente, às funções densidade e distribuição de probabilidade normais [104].

Seja a variável \bar{x} definida como o valor do logaritmo natural do módulo do ganho normalizado em relação à média e ao desvio padrão, ambos determinados estatisticamente. Com base no teorema fundamental da função densidade, a transformação de \bar{x} no parâmetro y , que será uniformemente quantizado, é dada por:

$$y = 0,5 + \frac{1}{\sqrt{2\pi}} \int_0^{\bar{x}} e^{-z^2/2} dz \quad (5.1)$$

A equação (5.1) pode ser resolvida através de algum método de integração numérica, como o método de Simpson, por exemplo. A substituição da integral por um somatório, com alocação dinâmica do número de parcelas segundo um limiar de erro desejado, mostrou-se mais eficiente. Desta forma:

$$y \cong 0,5 + \frac{1}{\sqrt{2 \cdot \pi}} \sum_{n=0}^P \left[\frac{\bar{x}^{2 \cdot n+1}}{(2 \cdot n+1) \cdot (-2)^n \cdot n!} \right] \quad (5.2)$$

onde P é o número de parcelas para um erro pré-definido.

O valor de P para um determinado erro no cálculo de y é uma função de \bar{x} . Neste caso, fixando-se a grandeza do erro desejado, é possível determinar o número mínimo de parcelas do somatório em (5.2) como uma função de \bar{x} . A equação (5.3) apresenta uma aproximação linear para P em função de \bar{x} , considerando um erro da ordem de 10^{-8} na determinação de y.

$$P = \text{INT}(7,728998 \cdot |\bar{x}| + 2,698350) \quad (5.3)$$

Os valores obtidos para y em (5.2) são então uniformemente quantizados entre os limites de 0,001349898 e 0,998650101, o que equivale a considerar somente valores contidos no intervalo de $\pm 3 \cdot \sigma$ na distribuição normal, sendo obtido então o parâmetro y_q . A transformação de y_q em valores quantizados do logaritmo natural do módulo do ganho consiste em, dado um valor de $y = y_q$ na equação (5.2), determinar-se o valor correspondente de \bar{x} . Tal problema deve ser solucionado de forma numérica e para tanto, foi aplicado o método da secante ou método da falsa posição [105].

O valor de \bar{x} procurado (\bar{x}_∞) é a raiz real da equação $g(\bar{x})$ dada a seguir:

$$g(\bar{x}) = 0,5 + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^P \left[\frac{\bar{x}^{2n+1}}{(2n+1) \cdot (-2)^n \cdot n!} \right] - y_q \quad (5.4)$$

A utilização do método da falsa posição para determinar \bar{x} , tal que $g(\bar{x}) \cong 0$, conduz a:

$$\bar{x}_{i+1} = \bar{x}_i \cdot (y_q - 0,5) / \left[\frac{1}{\sqrt{2\pi}} \sum_{n=0}^P \frac{\bar{x}_i^{2n+1}}{(2n+1) \cdot (-2)^n \cdot n!} \right] \quad (5.5)$$

onde:

\bar{x}_i é o valor atual de \bar{x} ;

\bar{x}_{i+1} é o novo valor de \bar{x} .

A equação (5.5) é calculada iterativamente até que $(\bar{x}_{i+1} - \bar{x}_i)$ atinja um limiar de erro pré-estabelecido.

O método proposto não apresentou problemas de convergência, entretanto a velocidade de convergência dependente do valor de \bar{x} , e diminui consideravelmente para valores distantes da origem. Outro fator importante na velocidade de determinação da solução de (5.4) é a escolha do valor inicial de \bar{x} , \bar{x}_0 . A função representada por (5.6) tem forma próxima da função $\bar{x} = f(y_q)$ desejada e pode ser usada para obter a aproximação \bar{x}_0 .

$$\bar{x}_0 = \tanh^{-1}(2 \cdot y_q - 1) \quad (5.6)$$

Os valores de \bar{x} obtidos de (5.5) devem ser finalmente desnormalizados em relação à média e ao desvio padrão, resultando nos valores quantizados do logaritmo do módulo do ganho.

A aplicação direta do método de quantização proposto, com as equações

apresentadas, aumenta consideravelmente a complexidade computacional da etapa de análise do codificador CELP, já que o procedimento deve ser realizado para cada candidata do "codebook". Um procedimento prático, que não implica o aumento da complexidade computacional em relação a outros métodos, consiste em distribuir uniformemente os níveis de quantização no domínio da variável y e obter através da equação (5.5), os respectivos níveis no domínio de \bar{x} , criando uma tabela de valores para quantização do ganho. A tabela de quantização pode ainda ser projetada de maneira a incluir a operação $\ln|g|$, diminuindo ainda mais a complexidade computacional no procedimento de quantização do ganho.

5.3 - ANÁLISE DO DESEMPENHO DOS CODIFICADORES CELP E CMSEC COM QUANTIZAÇÃO ESTATÍSTICA DOS GANHOS

Para aplicação do procedimento de quantização estatística foram determinadas as médias e os desvios padrão do logaritmo natural do módulo de cada ganho nos codificadores CELP e CMSEC, a partir das ocorrências destes parâmetros para os arquivos relacionados no Apêndice C. Os limites superior e inferior foram fixados em $+3.\sigma$ e $-3.\sigma$, exceto nos casos onde nenhuma ocorrência atingiu estes extremos, caso em que o limite foi considerado como a ocorrência mais próxima de $\pm 3.\sigma$.

A Tabela 5.1 apresenta os valores levantados estatisticamente e utilizados no procedimento de quantização estatística do codificador CELP.

A Tabela 5.2 apresenta os resultados objetivos obtidos com a aplicação do procedimento de quantização estatística do ganho no codificador CELP convencional para diversos números de bits (incluindo 1 bit para representação do sinal do ganho), bem como os resultados obtidos com o procedimento de quantização uniforme do logaritmo do módulo do ganho. Pode ser verificado o

aumento no tempo de processamento, considerável em alguns casos, provocado pela utilização do procedimento de quantização estatística com as equações (5.2) e (5.5).

Tabela 5.1 - Valores utilizados para quantização estatística do ganho no codificador CELP.

Logarítmo natural de $ g $			
μ	σ	y máximo	y mínimo
2,386776	1,007282	0,998650101	0,001349898

Tabela 5.2 - Resultados objetivos em função do procedimento de quantização do ganho no codificador CELP com 1024 "codewords".

Arquivo	Bits/g	Taxa(bps)	RSRSEG(dB)	Tp(min.)	Quantização
AM/AF	5	7200	12,29/11,61	42/45	uniforme
AM/AF	4	7000	12,09/11,44	42/45	uniforme
AM/AF	3	6800	11,65/10,95	42/45	uniforme
AM/AF	2	6600	8,20/ 7,93	42/45	uniforme
AM/AF	1	6400	6,72/ 6,53	42/45	uniforme
AM/AF	5	7200	12,25/11,60	105/98	estatística
AM/AF	4	7000	12,20/11,51	86/85	estatística
AM/AF	3	6800	11,89/11,30	70/72	estatística
AM/AF	2	6600	11,33/10,75	58/61	estatística
AM/AF	1	6400	10,00/ 9,88	46/49	estatsitica

● Análise subjetiva informal: Com a utilização de 5 bits na quantização do ganho, ambos os procedimentos implicaram resultados semelhantes, sendo que o procedimento estatístico apresentou-se sutilmente menos ruidoso. Com a utilização de 4 bits na quantização do ganho, o procedimento estatístico apresentou desempenho equivalente ao procedimento uniforme com 5 bits. Já com a utilização de 3 bits, a quantização estatística resultou em uma ligeira degradação em relação ao procedimento de quantização uniforme com 5 bits,

somente percebida na audição com fones de ouvido. Os resultados obtidos com quantização estatística utilizando 1 ou 2 bits e com quantização uniforme utilizando 3 bits ou menos, apresentaram degradações mais acentuadas sendo, em todos os casos, os resultados obtidos com o procedimento de quantização estatística sensivelmente superiores.

Tabela 5.3 - Valores utilizados para quantização estatística dos ganhos nos codificadores CMSEC de 2, 3, 5 e 8 ganhos.

Ganho	$\ln g_1 $	2 Ganhos	3 Ganhos	5 Ganhos	8 Ganhos
g_1	μ	1,758742	1,552327	1,474860	1,045625
	σ	1,340061	1,384746	1,414486	1,419756
	y máx.	0,998650101	0,998650101	0,998650101	0,998650101
	y mín.	0,001349898	0,001349898	0,001349898	0,001349898
g_2	μ	1,775313	1,478924	1,504981	1,433157
	σ	1,325729	1,393722	1,408700	1,436520
	y máx.	0,998650101	0,997515328	0,997618394	0,997460218
	y mín.	0,001349898	0,001349898	0,001349898	0,001349898
g_3	μ	-	1,567152	1,419023	0,979810
	σ	-	1,382589	1,425916	1,338157
	y máx.	-	0,998650101	0,998650101	0,998650101
	y mín.	-	0,001349898	0,001349898	0,001349898
g_4	μ	-	-	1,403678	0,943839
	σ	-	-	1,420102	1,405216
	y máx.	-	-	0,998650101	0,997208565
	y mín.	-	-	0,001349898	0,001349898
g_5	μ	-	-	1,474803	1,451960
	σ	-	-	1,403557	1,405247
	y máx.	-	-	0,998225327	0,997776179
	y mín.	-	-	0,001349898	0,001349898
g_6	μ	-	-	-	1,015357
	σ	-	-	-	1,408857
	y máx.	-	-	-	0,997289202
	y mín.	-	-	-	0,001349898
g_7	μ	-	-	-	1,450945
	σ	-	-	-	1,437108
	y máx.	-	-	-	0,997841528
	y mín.	-	-	-	0,001349898
g_8	μ	-	-	-	1,120323
	σ	-	-	-	1,459430
	y máx.	-	-	-	0,996985313
	y mín.	-	-	-	0,001349898

A Tabela 5.3 apresenta os valores levantados estatisticamente para a aplicação do procedimento de quantização estatística nos codificadores CMSEC de 2, 3, 5 e 8 ganhos.

A Tabela 5.4 apresenta os resultados objetivos obtidos com a aplicação do procedimento de quantização estatística em várias configurações do codificador CMSEC.

Tabela 5.4 - Resultados objetivos com a aplicação de quantização estatística dos ganhos no codificador CMSEC.

Arquivo	Ganhos	Bits/g	Codeb.	Taxa(bps)	RSRSEG(dB)	Tp(min.)
AM/AF	2	4	256	7400	12,37/11,67	40/40
AM/AF	3	4	64	7800	12,31/11,56	19/19
AM/AF	3	3	64	7200	11,80/11,20	18/18
AM/AF	5	4	128	9600	14,05/13,10	46/48
AM/AF	8	5	2	12400	13,19/12,14	8/ 8
AM/AF	8	4	2	10800	12,63/11,72	8/ 8
AM/AF	8	3	2	9200	11,47/10,87	8/ 8

• Análise subjetiva informal: A utilização de quantização estatística resultou em melhoras sensíveis nos resultados finais, sobretudo naqueles onde os ganhos foram quantizados com 3 bits. Em 7400 bps, o resultado foi semelhante ao do CELP a 7200 bps, enquanto que em 7200 bps e 7800 bps foi percebido um ligeiro aumento do ruído de fundo na audição com fones de ouvido. Já em 9600 bps, o resultado foi semelhante ao do CELP à mesma taxa, tendo sido ligeiramente menos "abafado" que o último na audição com fones. Os resultados obtidos em 10800 bps e 12400 bps apresentaram-se um pouco mais ruidosos, mas ainda assim de boa qualidade, sendo que o primeiro em pouco se diferenciou do CELP à 7200 bps. Para a taxa de 9200 bps o arquivo AM apresentou uma degradação maior na audição com fones. Na audição através das caixas

acústicas, todas as simulações apresentaram bons resultados, com diferenças muito pequenas entre si.

A partir do codificador CMSEC de 5 ganhos e taxa de 9600 bps, foi simulado o comportamento do mesmo codificador a 4800 bps, através do aumento dos quadros e subquadros, sendo os resultados objetivos apresentados na Tabela 5.5.

Tabela 5.5 - Resultados objetivos para o codificador CMSEC a 4800 bps com quantização estatística dos ganhos.

Arquivo	Ganhos	Bits/g	Codeb.	RSRSEG(dB)	Tempo(min.)	Observação
AM/AF	5	4	128	9,63/ 9,28	36/38	s/pós-filtr.
AM/AF	5	4	128	9,39/ 8,90	36/38	c/pós-filtr.

• Análise subjetiva informal: Novamente os resultados obtidos com quantização estatística foram superiores em relação ao procedimento de quantização uniforme, embora ainda ligeiramente inferiores aqueles obtidos com o CELP convencional a 4800 bps. A utilização de pós-filtragem mostrou-se vantajosa mais uma vez.

5.4 - CONCLUSÕES

O procedimento de quantização estatística proposto, consistindo na transformação dos valores do logaritmo natural do módulo do ganho em uma variável com distribuição de probabilidade uniforme e limitada entre 0 e 1, apresentou bons resultados, tanto no codificador CELP, como no CMSEC.

A utilização do procedimento de quantização estatística do ganho no codificador CELP resultou em um ganho subjetivo de aproximadamente 1 bit, ou

na melhoria da qualidade para a mesma taxa, em relação ao procedimento de quantização uniforme do logaritmo do módulo do ganho. Os resultados com 5 e 4 bits na quantização uniforme podem ser considerados bons, enquanto que com o procedimento de quantização estatística, a utilização de até 3 bits, incluindo o bit de sinal, conduziu a resultados aceitáveis.

O desempenho do codificador CMSEC foi sensivelmente melhorado com a utilização da quantização estatística dos ganhos, sobretudo naquelas configurações onde foram utilizados 3 bits na quantização de cada ganho.

A implementação do procedimento proposto diretamente através da utilização das equações de transformação (5.2) e (5.5) acarreta um considerável aumento da complexidade computacional, já que o procedimento é executado para cada candidata do "codebook". Já a utilização de uma tabela de quantização, gerada com auxílio da equação (5.5), não acarreta aumento da complexidade computacional em relação a outros procedimentos de quantização.

Apesar de ter apresentado resultados iniciais vantajosos em relação ao procedimento de quantização uniforme do logaritmo do módulo do ganho, a comparação do método proposto com outros métodos de quantização, tais como a utilização do algoritmo LBG, faz-se necessária com o objetivo de melhor avaliar o seu desempenho.

CAPÍTULO 6

DETERMINAÇÃO RÁPIDA DO RETARDO
DO FILTRO PLA6.1 - INTRODUÇÃO

O procedimento de determinação dos parâmetros do filtro PLA demanda considerável esforço computacional, somente suplantado pela busca da melhor "codeword" no procedimento de análise do codificador CELP convencional. Dentre os métodos existentes para determinação do atraso M do filtro PLA, o procedimento exaustivo da busca da menor energia do erro ponderado para toda a faixa de variação de M é aquele que implica melhores resultados. A complexidade deste método, entretanto, faz com que, em algumas implementações, seja adotado o procedimento de determinação de M através da busca do máximo da função correlação cruzada normalizada do resíduo do filtro PCA ou do próprio sinal de voz. Este último procedimento também implica uma quantidade razoável de operações algébricas, já que necessita do cálculo de uma expressão do tipo de (3.9) para todos os valores possíveis de M .

Neste capítulo é proposto um método para a redução da complexidade computacional na determinação do atraso M do filtro PLA, denominado busca rápida de M . A nova técnica é baseada em um método apresentado por Singhal e Atal em [25], para determinação da localização dos pulsos na excitação do codificador multipulso ("search pruning"). O método consiste, basicamente, na concentração da busca de M em torno dos máximos de uma função correlação

cruzada ou em torno dos mínimos da função energia do erro ponderado, conforme o procedimento utilizado. O método proposto é aplicado nos codificadores CELP e CMSEC, com a finalidade de avaliar seu desempenho e de obter o melhor compromisso entre redução de complexidade computacional e manutenção da qualidade do sinal de voz sintetizado.

6.2 - PROCEDIMENTO DE BUSCA RÁPIDA DE M

O procedimento de determinação ótima de M implica a busca do valor mínimo da energia do erro ponderado em função de M , $E(M)$, para toda a faixa de variação deste parâmetro (ex. 20...147 amostras) ou, equivalentemente, do valor máximo do inverso da função energia do erro ponderado, $E^{-1}(M)$. A análise da função $E^{-1}(M)$ demonstra que a mesma possui picos bem distintos, sobretudo para os segmentos de voz não ruidosos. As Fig.6.1 e Fig.6.2 apresentam a forma da função $E^{-1}(M)$ para segmentos vozeados e não vozeados, respectivamente. Em vista desta característica e considerando ainda que a influência do filtro PLA é mais significativa nos segmentos vozeados, a utilização de um método que concentre a busca do máximo da função $E^{-1}(M)$ em torno de seus picos, pode conduzir à redução da complexidade computacional na determinação de M .

Em [25] é apresentado um procedimento para redução da complexidade computacional na alocação dos pulsos na excitação do codificador multipulso. Tal procedimento tem por finalidade a determinação do ponto de máximo de uma função redução de erro, e está baseada no fato de que a função pesquisada varia lentamente e tem somente alguns poucos picos. Uma maneira de obter o resultado desejado é através da amostragem da função redução do erro a intervalos cada vez mais menores, até que todos os picos tenham sido encontrados.

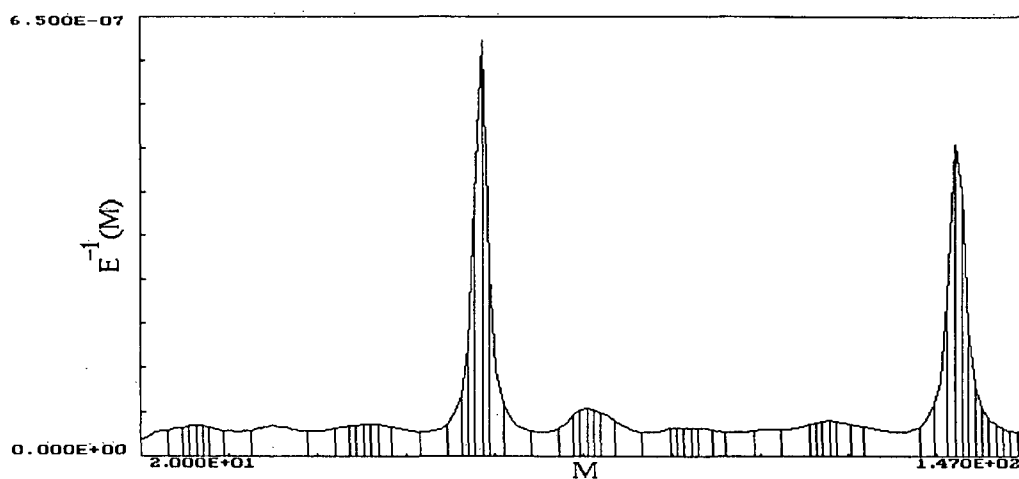


Fig.6.1 - Função $E^{-1}(M)$ para um segmento vozeado.

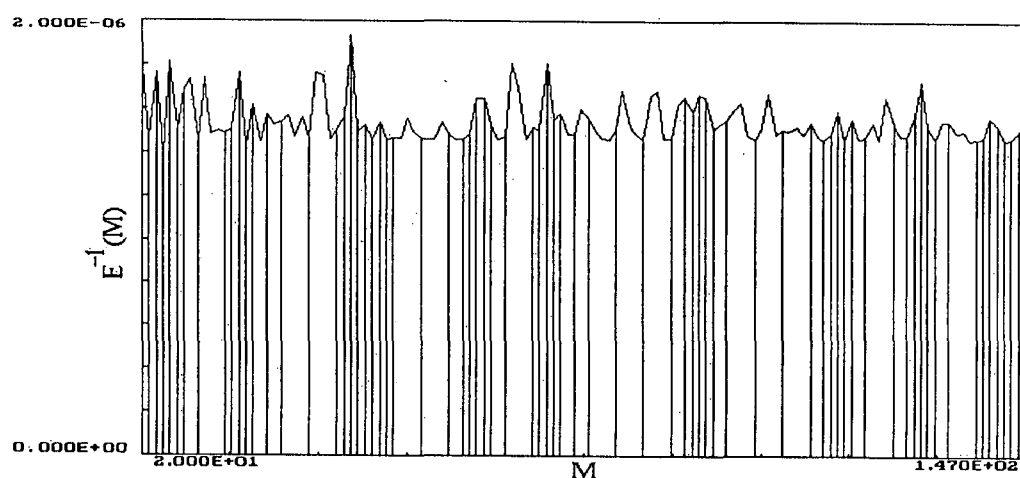


Fig.6.2 - Função $E^{-1}(M)$ para um segmento não vozeado.

Considerando uma faixa de valores $M_i \leq M \leq M_f$, onde M_i e M_f são, respectivamente, os valores mínimo e máximo possíveis para M , o procedimento para a determinação do máximo da função $E^{-1}(M)$ pode ser formalizado como a seguir:

- a) fazer $\Delta = \Delta_0$ (passo de amostragem inicial);

- b) calcular $E^{-1}(M)$ para os valores $M = M_1, \Delta+M_1, 2.\Delta+M_1, \dots, M_r$;
- c) para cada $M_0 \in \{M \mid E^{-1}(M) \geq E^{-1}(M \pm \Delta)\}$ calcular $E^{-1}(M_0 \pm \Delta/2)$;
- d) reduzir Δ pela metade ($\Delta_i = \Delta_{i-1}/2$);
- e) retornar ao passo c) se $\Delta \geq 1$.

Após a execução do último passo, o maior valor calculado para $E^{-1}(M)$ é considerado como o máximo da função e sua posição determina o valor de M procurado. A Fig.6.3 ilustra o procedimento descrito.

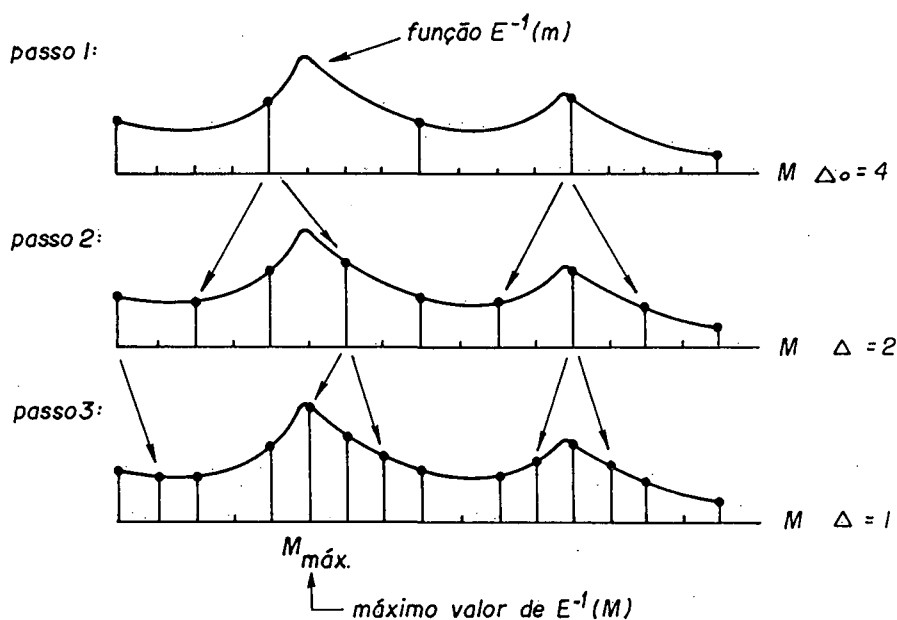


Fig.6.3 - Procedimento de busca rápida de M .

A redução de complexidade obtida com o algoritmo proposto é dependente do sinal, sendo possível entretanto a determinação experimental de um número de pontos a serem pesquisados na função $E^{-1}(M)$, que implique em perda mínima de picos com o procedimento proposto. Outro parâmetro a ser determinado experimentalmente é o valor de Δ_0 , que para o procedimento adotado deve ser uma potência de 2 e cujo aumento implica a redução da complexidade ao custo da

possível exclusão de alguns picos no procedimento de busca. A escolha de passo inicial $\Delta_0 = 1$ resulta na determinação do máximo global de $E^{-1}(M)$ em todas as situações, já que todos os possíveis valores de M são pesquisados. Para $\Delta_0 > 1$ não é garantida a determinação do máximo global de $E^{-1}(M)$. É esperado ainda um máximo na redução da complexidade para os segmentos vozeados ou que apresentem picos bem distintos.

O procedimento de busca rápida de M pode também ser aplicado em conjunto com o método de determinação de M através da procura do máximo da função correlação cruzada do resíduo do filtro PCA ou da função correlação cruzada do próprio sinal de voz.

6.3 - ANÁLISE DO DESEMPENHO DOS CODIFICADORES CELP E CMSEC COM O PROCEDIMENTO DE BUSCA RÁPIDA DE M

Com o objetivo de definir o parâmetro Δ_0 mais adequado para o procedimento de busca rápida de M , foram realizadas diversas simulações com o codificador CELP convencional operando a 7200 bps. A Tabela 6.1 apresenta os resultados obtidos.

Tabela 6.1 - Resultados objetivos em função de diversos Δ_0 para o codificador CELP a 7200 bps.

Arquivo	Δ_0	RSRSEG(dB)	Tp(min.)
AM/AF	1	12,29/11,61	42/45
AM/AF	2	12,27/11,61	40/42
AM/AF	4	12,23/11,53	39/41
AM/AF	8	12,25/11,49	38/41
AM/AF	16	11,98/11,16	38/40
AM/AF	32	11,66/11,06	38/40

• Análise subjetiva informal: As simulações utilizando $\Delta_0 = 16$ e $\Delta_0 = 32$ apresentaram degradações perceptíveis. As demais simulações apresentaram muito pouca diferença em relação ao procedimento de determinação exaustiva de M ($\Delta_0 = 1$). Em particular, a utilização de $\Delta_0 = 8$ apresentou o melhor compromisso entre qualidade da voz sintetizada e redução de complexidade computacional.

Foram também simulados os resultados do codificador CMSEC com 8 ganhos operando a 10800 bps e utilizando o procedimento de busca rápida de M. Os resultados objetivos são apresentados na Tabela 6.2.

Tabela 6.2 - Resultados objetivos para o codificador CMSEC a 10800 bps com o procedimento de busca rápida de M.

Arquivo	Δ_0	Ganhos	Bits/g	Codebook	RSRSEG(dB)	Tp(min.)
AM/AF	8	8	4	2	12,47/11,57	4/ 5

• Análise subjetiva informal: Os resultados apresentaram-se mais ruidosos do que aqueles obtidos com o CELP convencional a 7200 bps, não sendo porém de má qualidade. A utilização do procedimento de determinação rápida de M não degradou de forma perceptível o resultado.

O processamento de todos os arquivos mencionados no Apêndice C através do codificador CELP operando a 7200 bps, com busca rápida de M e $\Delta_0 = 8$, conduziu aos resultados apresentados na Tabela 6.3.

Pode ser observado que no pior caso, menos da metade (45,57 %) dos filtros PLA foram calculados para determinação do M sub-ótimo e que a relação sinal/ruído segmentada média caiu somente 0,07 dB em relação à utilização de $\Delta_0 = 1$. Tais resultados, aliados a qualidade subjetiva obtida, atestam a

validade do procedimento proposto para a redução de complexidade na determinação do filtro PLA.

Tabela 6.3 - Resultados objetivos para o codificador CELP a 7200 bps, com $\Delta_0 = 1$ e $\Delta_0 = 8$.

	RSRSEG(dB)		PLAs Calculados(%)
	$\Delta_0=1$	$\Delta_0=8$	$\Delta_0=8$
μ	14,48	14,41	44,16
σ	1,51	1,52	1,03
máx.	17,45	17,34	45,57
mín.	9,99	9,79	42,49

Finalmente, na Tabela 6.4 são apresentados os resultados subjetivos para os codificadores CELP e CMSEC com 5 ganhos e 128 "codewords", operando a 9600 bps e utilizando os procedimentos de quantização estatística e de busca rápida de M.

Tabela 6.4 - Resultados objetivos para os codificadores CELP e CMSEC a 9600 bps com quantização estatística e busca rápida de M.

Arquivo	Δ_0	Ganhos	Bits/g	Codebook	RSRSEG(dB)	Tp(min.)
AM/AF	8	1	5	1024	13,70/12,94	131/121
AM/AF	8	5	4	128	13,99/12,95	42/ 45

● Análise subjetiva informal: Ambos os resultados apresentaram muito boa qualidade, tendo a aplicação dos procedimentos de quantização estatística do

ganho e de busca rápida de M no codificador CELP convencional a 9600 bps, implicando resultados ligeiramente menos ruidosos. Já o codificador CMSEC, também utilizando os dois procedimentos mencionados, apresentou resultados um pouco menos "abafados" do que aqueles obtidos com o CELP convencional operando à mesma taxa, implicando ruído de fundo ligeiramente mais perceptível na audição com fones de ouvido.

6.4 - CONCLUSÕES

A utilização do procedimento de determinação rápida do retardo M do filtro PLA implicou em sensível redução da complexidade computacional para a determinação dos parâmetros deste filtro, sem degradar perceptivelmente a qualidade do sinal sintetizado.

Com $\Delta_0 = 8$, um levantamento estatístico revelou que menos de 50 % dos filtros PLA foram efetivamente calculados para determinação do melhor filtro, com a introdução insignificante de degradação tanto a nível objetivo como a nível subjetivo.

O problema da dependência da redução de complexidade em relação ao sinal a ser codificado, pode ser contornado através da limitação do número de pontos a serem calculados para determinação do ponto de máximo da função.

Em vista dos resultados obtidos, julgamos que a aplicação do procedimento de busca rápida do retardo M é um procedimento altamente recomendável na diminuição da complexidade computacional de qualquer tipo de codificador que utilize um filtro PLA com estrutura e forma de determinação semelhantes ao do codificador CELP apresentado.

CAPÍTULO 7

CONCLUSÕES

O estudo, implementação e análise do desempenho do codificador CELP convencional, bem como propostas para o aumento do desempenho quanto à relação taxa/qualidade e para a redução da complexidade computacional na codificação de sinais de voz foram abordadas neste trabalho.

O estudo do codificador CELP convencional resultou em configurações com taxas de transmissão entre 2400 bps e 9600 bps que podem ser utilizadas como padrões de referência para comparação do desempenho de novas estruturas de codificação paramétrica. O codificador CELP implementado conduziu a resultados de muito boa qualidade em 7200 bps e 9600 bps. Em 4800 bps, a qualidade pode ainda ser considerada boa, apesar da necessidade da utilização de pós-filtragem, enquanto que em 3600 bps e 2400 bps foram verificados níveis de degradação acentuados no sinal de voz sintetizado. A alta complexidade computacional, sobretudo na etapa de codificação da excitação, demonstrou ser a grande limitação do codificador CELP convencional com relação à sua aplicação em tempo real.

Foi apresentada uma estrutura de codificação paramétrica, denominada CMSEC, onde a excitação da cascata de filtros PLA e PCA é formada através da combinação linear de várias "codewords" tomadas de "codebooks" de tamanho reduzido, todas referenciadas por um mesmo índice a ser transmitido ao decodificador. As simulações realizadas demonstraram que, através de um compromisso adequado entre taxa de transmissão e qualidade do sinal

sintetizado, é possível obter-se configurações com complexidade computacional bastante reduzida em relação ao codificador CELP convencional. O principal problema verificado no codificador CMSEC diz respeito à quantização dos ganhos, já que a utilização do procedimento de quantização escalar destes parâmetros implicou na introdução de degradação relativamente maior do que aquela verificada na quantização do ganho no codificador CELP.

Com o objetivo de melhorar o procedimento de quantização escalar dos ganhos, tanto no codificador CELP como no CMSEC, foi proposto um procedimento que considera uma aproximação da função densidade de probabilidade do logaritmo natural do módulo do ganho. O denominado procedimento de quantização estatística do ganho implicou em melhoras objetivas e subjetivas na qualidade do sinal de voz sintetizado, tendo sido constatado um ganho subjetivo de praticamente 1 bit na quantização do ganho do codificador CELP em relação aos resultados obtidos com o procedimento de quantização uniforme do logaritmo natural do módulo do ganho. A qualidade dos resultados obtidos com o codificador CMSEC melhorou significativamente com a utilização do procedimento proposto, sobretudo quando foram utilizados 3 bits na quantização de cada ganho. A aplicação do método de quantização estatística do ganho através da utilização de uma tabela de quantização, não implica aumento da complexidade computacional em relação a outros métodos de quantização do ganho.

A redução da complexidade computacional na determinação dos parâmetros do filtro PLA também foi abordada. A adaptação de um método originalmente proposto para a determinação da posição dos pulsos da excitação do codificador multipulso, possibilitou uma considerável redução da complexidade computacional na determinação do parâmetro M do filtro PLA. O procedimento proposto conduziu a uma redução de mais de 50 % na complexidade computacional para a determinação dos coeficientes do filtro PLA, sem introduzir degradações

sensíveis no sinal sintetizado, tanto objetivamente como subjetivamente.

Com relação ao parâmetro de medida de qualidade utilizado na avaliação objetiva dos resultados, a relação sinal/ruído segmentada, pode ser concluído que a mesma deve ser tratada com restrições, já que não manteve sempre coerência com os resultados verificados nas avaliações subjetivas. Desta forma, a utilização da RSRSEG para a comparação de resultados obtidos em estudos diversos, onde foram utilizadas estruturas de codificação e segmentos de voz diferentes, pode não ter qualquer significado prático.

Em vista do conhecimento adquirido com a execução deste trabalho, podem ser citadas algumas sugestões para trabalhos futuros, relacionadas a seguir:

- estudo, implementação e análise de desempenho dos diversos tipos de vocoders apresentados, com o objetivo de verificar as vantagens e as limitações de cada um [1];
- estudo, implementação e análise de desempenho de formas alternativas do codificador CELP convencional (VSELP [2,55,106], VXC [56], etc.) e comparação com os resultados obtidos neste trabalho;
- aplicação de procedimentos de quantização vetorial aos parâmetros do codificador CELP e correspondente verificação da relação custo/benefício [90];
- análise do comportamento dos codificadores CELP e CMSEC com a introdução de erros de transmissão, codificação de sinais corrompidos por ruído, operação tandem, e codificação de sinais de outra natureza (ex. dados) [107];
- aplicação de procedimentos de quantização vetorial ao conjunto de ganhos do codificador CMSEC e verificação da correspondente relação custo/benefício [90];
- estudo de outras formas de quantização do ganho para comparação com o

procedimento de quantização estatística [90];

- tentativa de implementação de uma configuração do codificador CMSEC para operação em tempo real, com taxa final menor do que 10 kbps, utilizando processadores de sinais da linha TMS320 ou similares;
- estudo de um parâmetro objetivo de medida da qualidade do sinal de voz sintetizado que reflita mais fielmente os resultados subjetivos [108-12].

Para finalizar, considerando o aumento gradativo do desempenho e redução de custo dos processadores de sinais disponíveis, o estudo de estruturas complexas de codificação de sinais de voz, antes de implementação praticamente inviável, justifica-se plenamente em vista do benefício introduzido por estes procedimentos, tanto em sistemas de comunicação pública ou privada, como nas chamadas máquinas falantes.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] RABINER, Lawrence R. & SHCAFER, Ronald W. Digital Processing of Speech Signals. New Jersey, Prentice-Hall, 1978. 512p.
- [2] SPANIAS, Andreas S. & WU, Frank H. Speech Coding and Speech Recognition Technologies: A Review. In: INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, Singapura, Anais... Singapura, IEEE, 1991, p. 572-7.
- [3] LEVINSON, Stephen E. & ROE, David B. A Perspective on Speech Recognition. IEEE Communications Magazine, E.U.A., :28-34, Jan. 1990.
- [4] NAIK, Jayant M. Speaker Verification: A Tutorial. IEEE Communications Magazine, E.U.A., :42-48, Jan. 1990.
- [5] FLANAGAN, J. L. & COKER, C. H. & RABINER, L. R. & SCHAFER, R. W. & UMEDA, N. Synthetic Voices for Computers. IEEE Spectrum, E.U.A., :22-45, Out. 1970.
- [6] HART, J.'t. & NOOTEBOOM, S. G. & VOGTEN, L. L. M. & WILLENS, L. F. Manipulação de Sons da Fala. Ciência e Técnica, Brasil, 2(6):179-91, 1983.
- [7] SAGISAKA, Yoshinori. Speech Synthesis from Text. IEEE Communications Magazine, E.U.A., :35-41/55, Jan. 1990.

- [8] DECINA, Maurizio & MODENA, Giulio. CCITT Standards on Digital Speech Processing. IEEE Journal on Selected Areas in Communications, E.U.A., 6(2):227-34, Feb. 1988.
- [9] SCHAFER, Ronald W. & RABINER, Lawrence R. Digital Representations of Speech Signals. Proceedings of the IEEE, E.U.A., 63(4):662-77, Abr. 1975.
- [10] FLANAGAN, James L. & SCHROEDER, Manfred R. & ATAL, Bishnu S. & CROCHIERE, Ronald E. & JAYANT, Nuggehally S. & TRIBOLET, Jose M. Speech Coding. IEEE Transactions on Communications, E.U.A., 27(4):710-37, Abr. 1979.
- [11] JAYANT, N. S. & NOLL, Peter. Digital Coding of Waveforms - Principles and Applications to Speech and Video. New Jersey, Prentice-Hall, 1984. 688p.
- [12] ALLEN, Jonathan & HUNNICUTT, M. Sharon & KLATT, Dennis. From Text to Speech. New York, Cambridge University Press, 1987, 215p.
- [13] JAYANT, N. S. High-Quality Coding of Telephone Speech an Wideband Audio. IEEE Communications Magazine, E.U.A., 28(1):10-20, Jan. 1990.
- [14] KITAWAKI, Nobuhiko & NAGABUCHI, Hiromi. Quality Assessment of Speech Coding and Speech Synthesis Systems. IEEE Communications Magazine, E.U.A., 36-44, Out. 1988.

- [15] BONNET, Madeleine & MACCHI, Odile & JAIDANE-SAIDANE, Meriem. Theoretical Analysis of the ADPCM CCITT Algorithm. IEEE Transactions on Communications, E.U.A., 38(6):847-58, Jun. 1990.

- [16] ATAL, Bishnu S. & SCHROEDER, Manfred R. Predictive Coding of Speech Signals and Subjective Error Criteria. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 27(3):247-54, Jun. 1979.

- [17] ATAL, Bishnu S. & SCHROEDER, Manfred R. Adaptive Predictive Coding of Speech Signals. The Bell System Technical Journal, E.U.A., :1973-86, Out. 1970.

- [18] FANT, Gunnar. Acoustic Theory of Speech Production. Holanda, Mouton & Co., 1970. 328p.

- [19] ATAL, Bishnu S. & HANAUER, Suzanne L. Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. The Journal of the Acoustical Society of America, E.U.A., 50(2/2):637-55, Ago. 1971.

- [20] ROCHA, Luis F. Predicción Lineal Aplicada a Señales de Voz. Revista Telegrafica Electrónica, Argentina, 799:962-5/969, Set. 1979 (Parte 1), 800:1139-41/1148, Out 1979 (Parte 2).

- [21] MCCREE, Alan V. & BARNWELL III, Thomas P. A New Mixed Excitation LPC Vocoder. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 593-6.

- [22] ROSE, Richard C. & BARNWELL III, Thomas P. Design and Performance of an Analysis-by-Synthesis Class of Predictive Speech Coders. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 38(9):1489-503, Set. 1990.

- [23] YAMAMOTO, José S. & VIOLARO, Fábio. Investigation on Vector Quantized Multipulse Excitation with Fixed Pattern of Positions. In: INTERNATIONAL TELECOMMUNICATIONS SYMPOSIUM, Rio de Janeiro, 1990. Anais... Rio de Janeiro, SBT/IEEE, 1990. p. 191-5.

- [24] ATAL, Bishnu S. High-Quality Speech at Low Bit Rates: Multi-Pulse and Stochastically Excited Linear Predictive Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 1681-684.

- [25] SINGHAL, Sharad & ATAL, Bishnu S. Amplitude Optimization and Pitch Prediction in Multipulse Coders. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 37(3):317-27, Mar. 1989.

- [26] BEROUTI, M. & JACHNER, J. & MERMELSTEIN P. Reducing Signal Delay in Multipulse Coding at 16 kb/s. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 3043-6.

- [27] CUSTÓDIO, Ricardo F. Codificadores Paramétricos de Sinais de Voz com Excitação Multi-Pulso. Dissertação de Mestrado, Florianópolis, UFSC, 1990. 90p.

- [22] ROSE, Richard C. & BARNWELL III, Thomas P. Design and Performance of an Analysis-by-Synthesis Class of Predictive Speech Coders. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 38(9):1489-503, Set. 1990.

- [23] YAMAMOTO, José S. & VIOLARO, Fábio. Investigation on Vector Quantized Multipulse Excitation with Fixed Pattern of Positions. In: INTERNATIONAL TELECOMMUNICATIONS SYMPOSIUM, Rio de Janeiro, 1990. Anais... Rio de Janeiro, SBT/IEEE, 1990. p. 191-5.

- [24] ATAL, Bishnu S. High-Quality Speech at Low Bit Rates: Multi-Pulse and Stochastically Excited Linear Predictive Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 1681-684.

- [25] SINGHAL, Sharad & ATAL, Bishnu S. Amplitude Optimization and Pitch Prediction in Multipulse Coders. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 37(3):317-27, Mar. 1989.

- [26] BEROUTI, M. & JACHNER, J. & MERMELSTEIN P. Reducing Signal Delay in Multipulse Coding at 16 kb/s. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 3043-6.

- [27] CUSTÓDIO, Ricardo F. Codificadores Paramétricos de Sinais de Voz com Excitação Multi-Pulso. Dissertação de Mestrado, Florianópolis, UFSC, 1990. 90p.

- [28] ROSE, Richard C. & BARNWELL III, Thomas P. The Self Excited Vocoder - An Alternate Approach to Toll Quality at 4800 bps. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 453-6.

- [29] ROSE, R. C. & BARNWELL III, Thomas P. Quality Comparison of Low Complexity 4800 bps Self Excited and Code Excited Vocoders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1987. Anais... Tokyo, IEEE, 1987. p. 1637-40.

- [30] KEMP, David P. & SUEDA, Retha A., & TREMAIN, Thomas E. An Evaluation of 4800 bps Voice Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Glasgow, 1989. Anais... Glasgow, IEEE, 1989. p. 200-3.

- [31] HERNÁNDEZ-GÓMEZ, L. A. & CASAJÚS-QUIRÓS, F. J. & GARCÍA-MATEO, C. & ORTEGA-GARCÍA, J. Real-Time Implementation and Evaluation of Variable Rate CELP Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 585-8.

- [32] ANDREOTTI, F. G. & MAIORANO, V. & VETRANO, L. & ZANONI F. Real-Time Implementation of a High Quality CELP Codec for Digital Mobile Radio System. INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, Singapura, 1991, Anais... Singapura, IEEE, 1991. p. 324-7.

- [33] CASAJÚS-QUIRÓS, F. J. & HERNÁNDEZ-GÓMEZ, L. A. & GARCIA-MATEO, C.

- Analysis and Quantization Procedures for a Real-Time Implementation of a 4.8 kb/s CELP Coder. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 609-12.
- [34] SCHROEDER, Manfred R. & ATAL, Bishnu S. Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tampa, 1985. Anais... Tampa, IEEE, 1985. p. 937-40.
- [35] SCHROEDER, Manfred R. & ATAL, Bishnu S. Speech Coding Using Efficient Block Codes. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Paris, 1982. Anais... Paris, IEEE, 1982. p. 1668-71.
- [36] OZAWA, Kazunori & MIYANO, Toshiki. 4kb/s Improved CELP Coder with Efficient Vector Quantization. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 213-6.
- [37] LINDE, Yoseph & BUZO, Andrés & GRAY, Robert M. An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications, E.U.A., 28(1):84-95, Jan. 1980.
- [38] SPERANZINI, Irvando L. Codificadores Paramétricos de Sinal de Voz com Excitação Codificada e Filtragem Adaptável. Dissertação de Mestrado, Florianópolis, UFSC, 1991. 107p.

- [39] DAVIDSON, Grant & GERSHO, Allen. Complexity Reduction Methods for Vector Excitation Coding. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 3055-8.

- [40] GALAND, C. & MENEZ, J. & ROSSO, M. Complexity Reduction of CELP Coders. In: INTERNATIONAL TELECOMMUNICATIONS SYMPOSIUM, Rio de Janeiro, 1990. Anais... Rio de Janeiro, SBT/IEEE, 1990. p. 181-5.

- [41] MENEZ, J. & GALAND, C. & ROSSO, M. & BOTTAU, F. Adaptive Code Excited Linear Predictive Coder (ACELP). In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Glasgow, 1989. Anais... Glasgow, IEEE, 1989. p. 132-5.

- [42] GIBSON, Jerry D. & CHEONG, Yoon C. & WOO, Hong C. & CHANG, Wen-Whei. A Comparison of Backward Adaptive Prediction Algorithms in Low Delay Speech Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 237-40.

- [43] CHEN, Juin-Hwey & LIN, Yen-Chun & COX, Richard V. A Fixed-Point 16 kb/s LD-CELP Algorithm. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 21-4.

- [44] CUPERMAN, Vladimir & GERSHO, Allen & PETTIGREW, Robert & YAO, Jey-Hsin. Low-Delay Vector Excitation Coding of Speech at 16 kb/s. IEEE

Transactions on Communications, E.U.A., 40(1):129-39, Jan. 1992.

- [45] MENEZ, J. & GALAND, C. & ROSSO, M. A 2 ms-Delay Adaptive Code Excited Linear Predictive Coder. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 457-60.

- [46] CHEN, Juin-Hwey. A Robust Low-Delay CELP Speech Coder at 16 kbit/s. In: GLOBECOM, Dallas, 1989. Anais... Dallas, IEEE, 1989. p. 1237-41.

- [47] CUPERMAN, Vladimir & GERSHO, Allen & PETTIGREW, Robert & SHYNK, John J. & YAO, Jey-Hsin. Backward Adaptation for Low Delay Vector Excitation Coding of Speech at 16 kbit/s. In: GLOBECOM, Dallas, 1989. Anais... Dallas, IEEE, 1989. p. 1242-6.

- [48] GRANZOW, W. & ATAL, Bishnu S. & PALIWAL, K. K. & SCHROETER, J. Speech Coding at 4 kb/s and Lower Using Single-Pulse and Stochastic Models of LPC Excitation. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 217-20.

- [49] KROON, Peter & DEPRETTERE, Ed F. & SLUYTER, Rob J. Regular-Pulse Excitation - A Novel Approach to Effective and Efficient Multipulse Coding of Speech. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 34(5):1054-63, Oct. 1986.

- [50] KROON, Peter & SLUYTER, Rob B. & DEPRETTERE, Ed. F. A Low Complexity

Regular Pulse Coding Scheme with a Reduced Transmission Delay. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 3083-6.

- [51] KIPPER, U. & REININGER, H. & WOLF, D. Improved CELP Coding Using Adaptive Excitation Codebooks. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 237-40.
- [52] AKAMINE, Masami & MISEKI, Kimio. Efficient Excitation Model for Low Bit Rate Speech Coding. INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, Singapura, 1991, Anais... Singapura, IEEE, 1991. p. 586-9.
- [53] AKAMINE, Masami & MISEKI, Kimio. CELP Coding with an Adaptive Density Pulse Excitation Model. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 29-32.
- [54] MARLOW, S. & BUGGY, B. Classified Vector Excitation Speech Coding. IEE Proceedings, Inglaterra, 136-1(5):325-7, Out. 1989.
- [55] GERSON, Ira A. & JASIUK, Mark A. Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kbps. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 461-4.
- [56] DAVIDSON, Grant & GERSHO, Allen. Multiple-Stage Vector Excitation Coding

of Speech Waveforms. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New York, 1988. Anais... New York, IEEE, 1988. p. 163-6.

- [57] LIN, X. & SALAMI, R. A. & STEELE, R. High Quality Audio Coding Using Analysis-by-Synthesis Technique. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 3617-20.
- [58] COPPERI, Maurizio & SERENO, Daniele. CELP Coding for High-Quality Speech at 8 kbit/s. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 1685-8.
- [59] ROY, Guylain & KABAL, Peter. Wideband CELP Speech Coding at 16 kbits/sec. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 17-20.
- [60] MAKHOUL, John. Linear Prediction: A Tutorial Review. Proceedings of the IEEE, E.U.A., 63(4):561-80, Abr. 1975.
- [61] MISEKI, Kimio & AKAMINE, Masami. Adaptive Bit-Allocation Between the Polo-Zero Synthesis Filter and Excitation in CELP. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 229-32.
- [62] HUANG, Z. & YANG, X. & ZHU, X. & KUH, A. Homomorphic Linear Predictive

Coding: A New Estimation Algorithm for All-Pole Speech Modelling. IEEE Proceedings, Inglaterra, 137-1(2):103-8, Abr. 1990.

- [63] KABAL, P. & MONCET, J.-L. & CHU, C. C. Synthesis Filter Optimization and Coding: Applications to CELP. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New York, 1988. Anais... New York, IEEE, 1988. p. 147-150.
- [64] DAVID, Sumam & RAMAMURTHI, Bhaskar. Two-Sided Filters for Frame-Based Prediction. IEEE Transactions on Signal Processing, E.U.A., 39(4):789-94, Abr. 1991.
- [65] BAÚ, Norberto & SEARA, Rui & BERMUDEZ, José C. M. Design of IIR Digital Filters from Time-Domain Specifications. INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, Singapura, 1991, Anais... Singapura, IEEE, 1991. p. 2443-6.
- [66] BAÚ, Norberto. Estrutura Mista FIR-IIR Adaptável para Identificação em Tempo Real de Sistemas com Resposta ao Impulso de Longa Duração. Dissertação de Mestrado, Florianópolis, UFSC, 1991. 78p.
- [67] BLANKINSHIP, W. A. Note on Computing Autocorrelations. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 22(1):76-7, Fev. 1974.
- [68] KENDALL, Willian B. A New Algorithm for Computing Correlations. IEEE Transactions on Computers, E.U.A., 23(1):88-90, Jan. 1974.

- [69] BARNWELL III, Thomas P. Recursive Windowing for Generating Autocorrelation Coefficients for LPC Analysis. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 29(5):1062-6, Out. 1981.

- [70] CELLARIO, Luca & FERRARIS, Giuseppe & SERENO, Daniele. A 2 ms Delay CELP Coder. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Glasgow, 1989. Anais... Glasgow, IEEE, 1989. p. 73-6.

- [71] ATAL, Bishnu S. Predictive Coding of Speech at Low Bit Rates, IEEE Transactions on Communications, E.U.A., 30(4):600-14, Abr. 1982.

- [72] ATAL, Bishnu S. On Determining Partial Correlation Coefficients by the Covariance Method of Linear Prediction. Journal of Acoustical Society of America, E.U.A., 62(suplemento 1):S64, 1977.

- [73] TRANCOSO, Isabel M. & ATAL, Bishnu S. Efficient Search Procedures for Selecting the Optimum Innovation in Stochastic Coders. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 38(3):385-96, Mar. 1990.

- [74] KROON, Peter & ATAL, Bishnu S. Pitch Predictors with High Temporal Resolution. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 661-4.

- [75] MARQUES, J. S. & TRANCOSO, Isabel M. & Tribolet, José M. & ALMEIDA, L.

- B. Improved Pitch Prediction with Fractional Delays in CELP Coding. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 665-8.
- [76] RAMACHANDRAN, Ravi P. & KABAL, Peter. Stability and Performance Analysis of Pitch Filters in Speech Coders. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 35(7):937-46, Jul. 1987.
- [77] KROON, Peter & DEPRETTERE, Ed F. A Class of Analysis-by-Synthesis Predictive Coders for High Quality Speech Coding at Rates Between 4.8 and 16 kbits/s. IEEE Journal on Selected Areas in Communications, E.U.A., 6(2):353-63, Fev. 1988.
- [78] KLEIJN, W. Bastiaan & KRASINSKI, Daniel J. & KETCHUM, Richard H. Fast Methods for the CELP Speech Coding Algorithm. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 38(8):1330-42, Ago. 1990.
- [79] RAMACHANDRAN, Ravi P. & KABAL, Peter. Pitch Prediction Filters in Speech Coding. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 37(4):467-78, Abr. 1989.
- [80] TRANCOSO, Isabel M. & ATAL, Bishnu S. Efficient Procedures for Finding the Optimum Innovation in Stochastic Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 2375-8.

- [81] LE GUYADER, A. & DI FRANCESCO, R. & LAMBLIN, C. Derivation of Efficient CELP Coding Algorithms Using the Z-Transform Approach. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 209-12.
- [82] LEE, J. I. & UN, C. K. On Reducing Computational Complexity of Codebook Search in CELP Coding. IEEE Transactions on Communications, E.U.A., 38(11):1935-7, Nov. 1990.
- [83] HERNÁNDEZ-GÓMEZ, L. A. & CASAJÚS-QUIRÓS, F. J. & FIGUEIRAS-VIDAL, A. R. & GARCÍA-GÓMEZ, R. Reducing Complexity on a Code-Excited Linear Predictor. In: Signal Processing III: Theories and Applications. Holanda, Elsevier Science Publishers B. V., 1986. p. 481-3.
- [84] HERNÁNDEZ-GÓMEZ, L. A. & CASAJÚS-QUIRÓS, F. J. & FIGUEIRAS-VIDAL, A. R. & GARCÍA-GÓMEZ, R. On the Behaviour of Reduced Complexity Code-Excited Linear Prediction (CELP). In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1986. Anais... Tokyo, IEEE, 1986. p. 469-72.
- [85] ADOUL, J-P. & MABILLEAU, P. & DELPRAT, M. & MORISSETTE, S. Fast CELP Coding Based on Algebraic Codes. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1987. Anais... Tokyo, IEEE, 1987. p. 1957-60.
- [86] ADOUL, J-P. & LAMBLIN, C. A Comparison of Some Algebraic Structures for CELP Coding of Speech. In: INTERNATIONAL CONFERENCE ON ACOUSTICS,

SPEECH, AND SIGNAL PROCESSING, Tokyo, 1987. Anais... Tokyo, IEEE, 1987. p. 1953-6.

- [87] LAFLAMME, C. & ADOUL, J.-P. & SU, H. Y. & MORISSETTE, S. On Reducing Computational Complexity of Codebook Search in CELP Coder Through the Use of Algebraic Codes. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New Mexico, 1990. Anais... New Mexico, IEEE, 1990. p. 177-80.
- [88] SCHROEDER, Manfred R. & ATAL, Bishnu S. Stochastic Coding of Speech Signals at Very Low Bit Rates: The Importance of Speech Perception. Speech Communication, Holanda, 4(1-3):155-62, Ago. 1985.
- [89] CHEN, Juin-Hwey & GERSHO, Allen. Real-Time Vector APC Speech Coding at 4800 bps with Adaptive Postfiltering. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1987. Anais... Tokyo, IEEE, 1987. p. 2185-8.
- [90] GRAY, Robert M. Vector Quantization. IEEE ASSP Magazine, E.U.A., :4-29, Abr. 1984.
- [91] ATAL, Bishnu S. Stochastic Gaussian Model for Low-Bit Rate Coding of LPC Area Parameters. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1987. Anais... Tokyo, IEEE, 1987. p. 2404-7.
- [92] PALIWAL, K. K. & ATAL, Bishnu S. Efficient Vector Quantization of LPC

Parameters at 24 Bits/Frame. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 661-4.

- [93] ATAL, Bishnu S. & COX, Richard V. & KROON, Peter. Spectral Quantization and Interpolation for CELP Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Glasgow, 1989. Anais... Glasgow, IEEE, 1989. p. 69-72.
- [94] MARKEL, John D. & GRAY JR., Augustine H. Implementation and Comparison of Two Transformed Reflection Coefficient Scalar Quantization Methods. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 28(5):575-83, Oct. 1980.
- [95] SOONG, Frank K. & JUANG, Biing-Hwang. Line Spectrum Pair (LSP) and Speech Data Compression. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, San Diego, 1984. Anais... San Diego, IEEE, 1984. p. 1.10.1-.4.
- [96] SOONG, Frank K. & JUANG, Biing-Hwang. Optimal Quantization of LSP Parameters. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New York, 1988. Anais... New York, IEEE, 1988. p. 394-7.
- [97] KANG, George S. & FRANSEN, Lawrence J. Application of Line-Spectrum Pairs to Low-Bit-Rate Speech Encoders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tampa, 1985. Anais... Tampa,

IEEE, 1985. p. 244-7.

- [98] KONDOZ, A. M. & EVANS, B. G. CELP Base-Band Coder for High Quality Speech Coding at 9.6 to 2.4 kbps. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New York, 1988. Anais... New York, IEEE, 1988. p. 159-62.

- [99] WU, Yuhang & HANSEN, Henrik B. & LARSEN, Knud J. High Performance Coder: A Possible Candidate for the GSM Half-Rate System. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 613-6.

- [100] JAYANT, N. S. & CHEN, J.-H. Speech Coding with Time-Varying Bit Allocations to Excitation and LPC Parameters. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Glasgow, 1989. Anais... Glasgow, IEEE, 1989. p. 65-8.

- [101] KROON, Peter & ATAL, Bishnu S. Quantization Procedures for the Excitation in CELP Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 1987. Anais... Tokyo, IEEE, 1987. p. 1649-52.

- [102] KROON, Peter & ATAL, Bishnu S. Strategies for Improving the Performance of CELP Coders at Low Bit Rates. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, New York, 1988. Anais... New York, IEEE, 1988. p.151-4.

- [103] FERNANDES, Dênis & BERMUDEZ, José C. M. & SEARA, Rui. Quantização Estatística do Ganho em Codificadores CELP. In: 10 SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, Brasília, 1992. Anais... Brasília, UnB/TELEBRÁS, 1992. p. 271-6.
- [104] PAPOULIS, Athanasios. Probability, Random Variables and Stochastic Processes. New York, McGraw-Hill, 1965. 583p.
- [105] CARNAHAM, Brice & LUTHER, H. A. & WILKES, James O. Applied Numerical Methods. New York, John Wiley & Sons, 1969. 604p.
- [106] GERSON, Ira A. & JASIUK, Mark A. Techniques for Improving the Performance of CELP Type Speech Coders. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Toronto, 1991. Anais... Toronto, IEEE, 1991. p. 205-8.
- [107] COX, Richard V. & KLEIJN, W. B. & KROON, Peter. Robust CELP Coders for Noisy Backgrounds and Noisy Channels. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Glasgow, 1989. Anais... Glasgow, IEEE, 1989. p. 739-42.
- [108] GRAY, Robert M. & BUZO, Andrés & GRAY JR., Augustine H. & MATSUYAMA, Yasuo. Distortion Measures for Speech Processing. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 28(4):367-76, Ago. 1980.
- [109] DIMOLITSAS, S. Objective Speech Distortion Measures and their Relevance

to Speech Quality Assessments. IEE Proceedings, Inglaterra, 136-1(5):317-24, Out. 1989.

[110] KITAWAKI, Nobuhiko & NAGABUCHI, Hiromi & ITOH, Kenzo. Objective Quality Evaluation for Low-Bit-Rate Speech Coding Systems. IEEE Journal on Selected Areas in Communications, E.U.A., 6(2):242-8, Fev. 1988.

[111] HUGGINS, A. W. F. & NICKERSON, R. S. Speech Quality Evaluation Using "Phoneme-Specific" Sentences. Journal of Acoustical Society of America, E.U.A., 77(5):1896-906, Mai. 1985.

[112] DAUMER, William R. Subjective Evaluation of Several Efficient Speech Coders. IEEE Transactions on Communications, E.U.A., 30(4):655-62, Abr. 1982.

A P Ê N D I C E A

M É T O D O D E B L A N K I N S H I P P A R A O C Á L C U L O D E A U T O C O R R E L A Ç Õ E S

A determinação dos coeficientes do filtro PCA através do método da autocorrelação necessita do cálculo das autocorrelações $R(\tau)$ do sinal de voz para $0 \leq \tau \leq OP$, onde OP é a ordem do preditor de curto atraso.

A utilização do método proposto por Blankinship [67], descrito a seguir, permite uma redução de até aproximadamente 50% no número de multiplicações sem aumento no número de somas.

Seja $s(n)$ para $0 \leq n \leq N_q-1$ o sinal de voz ponderado por uma janela qualquer, para o qual deseja-se determinar a autocorrelação em τ :

$$R(\tau) = \sum_{i=0}^{N_q-1-\tau} s(i) \cdot s(i+\tau) \quad (A.1)$$

O procedimento usual requer $N_q - \tau$ multiplicações e igual número de adições. Uma redução significativa no número de multiplicações pode ser obtida considerando que para $\tau \neq 0$ aparecem em (A.1) termos utilizados duas vezes como multiplicandos e que portanto podem ser colocado em evidência.

Desta forma, $R(\tau)$ pode ser assim determinada:

$$R(\tau) = R_1(\tau) + R_2(\tau) \quad (A.2)$$

onde:

$$R_1(\tau) = \sum_{j=0}^{q-1} \sum_{i=0}^{\tau-1} s(2.j.\tau+i+\tau) \cdot [s(2.j.\tau+i) + s(2.j.\tau+i+2.\tau)] \quad (A.3)$$

$$R_2(\tau) = \sum_{i=0}^{b-1} s(2.q.\tau+i) \cdot s(2.q.\tau+i+\tau) \quad , \text{para } a=0 \quad (A.4)$$

$$R_2(\tau) = \sum_{i=0}^{b-1} s(2.q.\tau+i+\tau) \cdot [s(2.q.\tau+i) + s(2.q.\tau+i+2.\tau)] + \sum_{i=b}^{\tau-1} s(2.q.\tau+i) \cdot s(2.q.\tau+i+\tau) \quad , \text{para } a=1 \quad (A.5)$$

Os parâmetros q , a e b devem satisfazer a relação a seguir:

$$N_q - \tau = 2.q.\tau + a.\tau + b \quad , a = 0 \text{ ou } 1 \quad \text{e} \quad 0 \leq b \leq \tau \quad (A.6)$$

A determinação de q , a e b pode ser realizada utilizando o conjunto de equações a seguir:

$$q = \text{INT} [(N_q - \tau) / (2.\tau)] \quad , 0 < \tau < N_q \quad (A.7)$$

$$a = \text{INT} [(N_q - \tau - 2.q.\tau) / \tau] \quad , 0 < \tau < N_q \quad (A.8)$$

$$b = N_q - \tau \cdot (1 + 2.q + a) \quad (A.9)$$

onde:

$\text{INT}[x]$ é a parte inteira de x .

O número de multiplicações requeridas pelo método proposto é dado por:

$$q.\tau+b = (N_q-\tau+b)/2 \quad , \text{para } a=0 \quad (\text{A.10})$$

$$q.\tau+\tau = (N_q-b)/2 \quad , \text{para } a=1 \quad (\text{A.11})$$

A Fig. A.1 apresenta o número de multiplicações como função de τ para o método usual e para o método de Blankinship com $N_q = 240$ e $0 \leq \tau \leq 240$. Pode ser observado que para $\tau < N_q/2$ existe uma redução significativa no número de multiplicações, que atinge aproximadamente 50% em $\tau=1$.

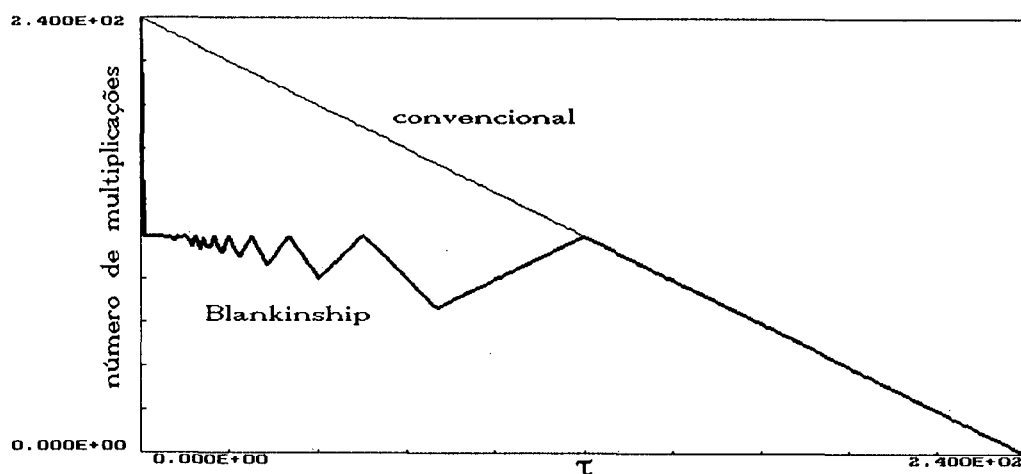


Fig.A.1 - Número de multiplicações no cálculo de $R(\tau)$.

As equações (A.2-5) e (A.7-9) implicam algumas multiplicações e somas adicionais para cada $R(\tau)$, o que não é significativo para $\tau \ll N_q$, além do que, a maior parte destas operações e o cálculo dos índices podem ser realizadas com aritmética inteira.

Para $\tau = 0$ não existe redução no número de multiplicações, e fazendo $q = a = 0$ e $b = N_q$, a autocorrelação $R(0)$ é determinada com base em (A.4), ou seja, $R(0) = R_2(0)$.

Os valores de q , a e b podem também ser tabelados em função de Nq e τ , o que não implica a necessidade de cálculo destes parâmetros, ao custo de redução da flexibilidade.

A P Ê N D I C E B

M É T O D O D E C H O L E S K Y P A R A R E S O L U Ç Ã O D E S I S T E M A S L I N E A R E S

A decomposição de Cholesky [1], também conhecida como Método da Raiz Quadrada, é um método de resolução de sistemas de equações lineares nos quais a matriz dos coeficientes apresenta a propriedade da simetria. Aproveitando-se deste fato, é possível reduzir a complexidade computacional na resolução do sistema em relação a outros métodos. Neste apêndice é apresentado o método da decomposição de Cholesky juntamente com um algoritmo para implementação prática do mesmo.

Seja o sistema de equações lineares dado a seguir:

$$\underline{\phi} \cdot \underline{\alpha} = \underline{\psi} \quad (\text{B.1})$$

onde $\underline{\phi}$ é uma matriz simétrica definida positiva com elementos ϕ_{ij} , $\underline{\alpha}$ é o vetor coluna que contém as incógnitas α_i e $\underline{\psi}$ é o vetor coluna com os termos independentes ψ_i .

A matriz $\underline{\phi}$ pode ser assim decomposta:

$$\underline{\phi} = \underline{V} \cdot \underline{D} \cdot \underline{V}^t \quad (\text{B.2})$$

onde \underline{V} é uma matriz triangular inferior com todos os elementos da diagonal principal iguais a 1 e \underline{D} é uma matriz diagonal.

Os elementos ϕ_{ij} podem ser determinados por:

$$\phi_{ij} = \sum_{k=1}^j V_{ik} \cdot D_k \cdot V_{jk}, \quad 1 \leq j \leq i \quad (\text{B.3})$$

Com base em (B.3), são determinados os elementos de \underline{V} e \underline{D} :

$$V_{ij} = [\phi_{ij} - \sum_{k=1}^{j-1} V_{ik} \cdot D_k \cdot V_{jk}] / D_j, \quad 2 \leq j \leq i \quad (\text{B.4})$$

$$V_{i1} = \phi_{i1} / D_1, \quad j = 1 \quad (\text{B.5})$$

$$D_j = \phi_{jj} - \sum_{k=1}^{j-1} V_{jk} \cdot D_k, \quad 2 \leq j \leq \text{OP} \quad (\text{B.6})$$

$$D_1 = \phi_{11} \quad (\text{B.7})$$

onde OP é a ordem do sistema.

Após a determinação de \underline{V} e \underline{D} , o vetor \underline{Y} pode ser calculado a partir de (B.1) e (B.2):

$$\underline{V} \cdot \underline{Y} = \underline{\psi} \quad (\text{B.8})$$

com

$$\underline{Y} = \underline{D} \cdot \underline{V}^t \cdot \underline{\alpha} \quad (\text{B.9})$$

onde

$$Y_i = \psi_i - \sum_{j=1}^{i-1} V_{ij} \cdot Y_j, \quad 2 \leq i \leq \text{OP} \quad (\text{B.10})$$

e

$$Y_1 = \psi_1 \quad (\text{B.11})$$

A determinação das incógnitas α_i é realizada através da seguinte relação:

$$\alpha_i = Y_i/D_i - \sum_{j=i+1}^{OP} V_{ji} \cdot \alpha_j, \quad OP-1 \geq i \geq 1 \quad (B.12)$$

com

$$\alpha_{OP} = Y_{OP}/D_{OP} \quad (B.13)$$

A seguir é apresentado um algoritmo para implementação do procedimento descrito:

1. $j=1$
2. $D_1 = \phi_{11}$
3. $V_{i1} = \phi_{i1}/D_1, \quad 2 \leq i \leq OP$
4. $j=j+1$
5. $D_j = \phi_{jj} - \sum_{k=1}^{j-1} V_{jk} \cdot D_k$
6. $V_{ij} = [\phi_{ij} - \sum_{k=1}^{j-1} V_{ik} \cdot D_k \cdot V_{jk}]/D_j, \quad j < i \leq OP$
7. volta ao passo 4 se $j < OP$
8. $Y_1 = \psi_1$
9. $Y_i = \psi_i - \sum_{j=1}^{i-1} V_{ij} \cdot Y_j, \quad 2 \leq i \leq OP$
10. $\alpha_{OP} = Y_{OP}/D_{OP}$
11. $\alpha_i = Y_i/D_i - \sum_{j=i+1}^{OP} V_{ji} \cdot \alpha_j, \quad OP-1 \geq i \geq 1$

A resolução de um sistema de equações lineares pela decomposição de Cholesky requer um número de multiplicações da ordem de $OP^3/2$ [60].

APÊNDICE C

ARQUIVOS COM OS SEGMENTOS DE VOZ
UTILIZADOS NAS SIMULAÇÕES

Para as simulações realizadas neste trabalho, foram utilizados os arquivos com segmentos de voz digitalizados relacionados na Tabela C.1, onde estão identificados o nome do locutor, idade, sexo, local de origem, posição na fita original ("tape deck" Gradiente DII) e número de amostras do sinal adquiridas. Todos os arquivos foram adquiridos em 12 bits, com quantização uniforme e frequência de amostragem de 8 kHz. As gravações originais foram realizadas em um "tape deck" cassete utilizando-se sistema Dolby B de redução de ruído e fita de metal, o que resultou em excelente qualidade para a aplicação desejada, com nível de ruído de fundo muito reduzido.

Todos os arquivos contém a mesma locução, que possui todos os fonemas da língua portuguesa com ocorrências proporcionais às suas frequências de ocorrência na fala cotidiana. A locução utilizada é descrita a seguir:

"Ontem meu filho fugiu de casa levando o seu cachorro, um boné preto, e ainda o dinheiro que ganhou da avó. O meu gostoso, quente e relaxante banho, mudou minha vontade de trabalhar, porque geralmente é um ótimo remédio no caso de fadiga."

Dos 39 arquivos relacionados, 22 contém segmentos de vozes masculinas e 17 contém segmentos de vozes femininas. Os arquivos 04 e 26 correspondem aos arquivos AM e AF, respectivamente, que foram utilizados na avaliação do desempenho das diversas configurações de codificadores testadas. No levantamento das ocorrências dos parâmetros para o estudo da quantização dos mesmos, foram utilizados 32 arquivos, 16 com vozes masculinas e 16 com vozes femininas, resultantes da exclusão dos arquivos 04, 26, 09, 11, 12, 13, e 27 da Tabela C.1.

Tabela C.1 - Arquivos com os segmentos de voz utilizados nas simulações.

N	Nome	Idade	Sexo	Origem	Contador	Amostras
01	Rui Seara	39	M	SC	0085-0124	117688
02	Elizena Stein	37	F	SC	0164-0200	109671
03	Rita Rocha	27	F	SC	0231-0268	117561
04	Sidnei Noceti Filho	38	M	SC	0300-0335	111684
05	Flávio Bressan	21	M	SC	0362-0393	101690
06	Paulo Sérgio dos Santos	21	M	SC	0420-0455	118818
07	Alexandre Ternes Behr	37	M	SC	0478-0519	140800
08	Volney Coelho Vincence	31	M	SP	0547-0582	122481
09	Rogério Koenen	21	M	SC	0606-0634	101346
10	Eloi Lenzi	25	M	PR	0662-0696	125473
11	Fábio Perez	22	M	PR	0727-0757	112035
12	José Gil Zipf	23	M	SC	0780-0811	118496
13	Itamar Loss	26	M	PR	0823-0856	127725
14	Dênis Fernandes	29	M	RS	0886-0920	134400
15	Irvando Luiz Speranzini	32	M	SC	0953-0981	113344
16	Eduardo Beck	25	M	SC	1009-1036	108778
17	Salete Machado	26	F	SC	1063-1089	108543
18	Elizabeth Ulbrich	37	F	SC	1108-1136	117030
19	Marlise Regina Matos	21	F	SC	1159-1185	109698
20	Marla Cristina Matos	22	F	SC	1210-1235	107060
21	Mary Ângela de Souza	30	F	SC	1270-1298	118985
22	Márcio Schneider	38	M	SC	1312-1339	125421
23	Cristina Marà	20	F	SC	1369-1397	123313
24	Fátima Cabrera	32	F	SP	1418-1444	118455
25	Lilian Terezinha Ribeiro	25	F	SP	1463-1488	112800
26	Vania Mattozo	26	F	SC	1507-1532	117855
27	Carlos Galup Montoro	38	M	Uruguai	1553-1583	144288
28	Edneu Mário Bizarre	21	M	SP	1610-1635	126336
29	Ângelo Mibielli	24	M	SC	1657-1680	113080
30	Elizete Maria Lourenço	22	F	SC	1699-1721	109920
31	Rosely Aparecida Peralta	29	F	PR	1743-1765	109975
32	Vitor Juliano De Negri	30	M	RS	1785-1808	112840
33	Simone Massulini Acosta	21	F	RS	1830-1851	108872
34	Gilberto Silva	22	M	SC	1871-1894	116808
35	Ricardo Takase	24	M	SP	1913-1935	112590
36	Rodrigo Bermudez	8	M	SC	1971-2006	178087
37	Eugenia Fernandes	69	F	RS	2026-2052	142000
38	Irene Fernandes	40	F	RS	2070-2102	172100
39	Crismara Janina da Rosa Santana	33	F	RS	2122-2144	122000

APÊNDICE D

TEMPO DE PROCESSAMENTO PARA
O CODIFICADOR CELP IMPLEMENTADO

A Tabela D.1 apresenta o tempo de processamento para diversas máquinas, expresso em minutos de processamento por segundo de voz, para o programa de simulação do codificador CELP operando a 7200 bps. A tabela apresenta ainda o tipo de otimização utilizada na compilação do programa em cada uma das máquinas.

O programa de simulação do codificador CELP convencional, assim como os demais programas utilizados neste trabalho, foram implementados em linguagem C, segundo o padrão ANSI, de forma a serem aceitos por qualquer máquina que possua compilador C compatível com este padrão.

O programa implementado simula simultaneamente as etapas de codificação e decodificação do codificador CELP, e foi construído com o objetivo principal de apresentar flexibilidade na variação dos parâmetros e procedimentos de determinação dos mesmos, o que compromete um pouco sua eficiência. A inclusão de sub-rotinas para cálculo de relação sinal/ruído e geração de relatórios sobre o processo, também contribuem para o aumento do tempo de processamento, podendo ser descartadas em uma implementação prática.

Tabela D.1 - Tempos de processamento do programa de simulação do codificador CELP convencional em 7200 bps para diversas máquinas.

Máquina	Otimização	Minutos/segundo de voz
IBM RISC 6000	0	3,47
Convex 210	02	3,62
Sun Sparc 1 ⁺ /IPC	fast	8,05
PC 486 50 MHz	speed	8,25
PC 486 33 MHz	speed	12,35
PC 286/80287 13 MHz	speed	284,60
PC V30/8087-II 8 MHz	speed	287,70